

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
DEPARTAMENTO ACADÊMICO DE INFORMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA**

**JULIO CESAR GONÇALVES**

**RECONHECIMENTO DE DÍGITOS EM IMAGENS DE MEDIDORES DE  
CONSUMO DE GÁS NATURAL UTILIZANDO TÉCNICAS DE VISÃO  
COMPUTACIONAL**

**DISSERTAÇÃO**

**CURITIBA**

**2016**

JULIO CESAR GONÇALVES

**RECONHECIMENTO DE DÍGITOS EM IMAGENS DE MEDIDORES DE  
CONSUMO DE GÁS NATURAL UTILIZANDO TÉCNICAS DE VISÃO  
COMPUTACIONAL**

Dissertação submetida ao Programa de Pós-Graduação em Computação Aplicada, da Universidade Tecnológica Federal do Paraná como requisito parcial para a obtenção do título de Mestre em Computação Aplicada.

Área de Concentração: Engenharia de Sistemas Computacionais.

Orientadora: Profa. Dra. Tania Mezzadri Centeno.

CURITIBA

2016

Dados Internacionais de Catalogação na Publicação

---

G635r  
2016  
Gonçalves, Julio Cesar  
Reconhecimento de dígitos em imagens de medidores de consumo de gás natural utilizando técnicas de visão computacional / Julio Cesar Gonçalves.-- 2016.  
102 f.: il.; 30 cm.

Texto em português, com resumo em inglês  
Disponível também via World Wide Web  
Dissertação (Mestrado) – Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Computação Aplicada, Curitiba, 2016  
Bibliografia: f. 97-102

1. Gás natural. 2. Reconhecimento ótico de caracteres. 3. Redes neurais (Computação). 4. Máquinas de vetores de suporte. 5. Algoritmos. 6. Computação – Dissertações. I. Centeno, Tania Mezzadri, orient. II. Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Computação Aplicada, inst. III. Título.

---

CDD: Ed. 22 -- 621.39

Biblioteca Central da UTFPR, Câmpus Curitiba

## ATA DE DEFESA DE DISSERTAÇÃO DE MESTRADO Nº 49

Aos 15 dias do mês de setembro de 2016 realizou-se na sala B 204 a sessão pública de Defesa da Dissertação de Mestrado intitulada "Reconhecimento de dígitos em imagens de medidores de consumo de gás natural utilizando técnicas de visão computacional", apresentado pelo aluno **Julio Cesar Gonçalves** como requisito parcial para a obtenção do título de Mestre em Computação Aplicada, na área de concentração "Engenharia de Sistemas Computacionais", linha de pesquisa "Sistemas de Informação".

### Constituição da Banca Examinadora:

Prof<sup>a</sup>. Dr<sup>a</sup>. Tania Mezzadri Centeno (Presidente) UTFPR \_\_\_\_\_

Prof<sup>o</sup>. Dr<sup>o</sup>. João Alberto Fabro UTFPR \_\_\_\_\_

Prof<sup>o</sup>. Dr<sup>o</sup>. Lucas Ferrari de Oliveira UFPR \_\_\_\_\_

Em conformidade com os regulamentos do Programa de Pós-Graduação em Computação aplicada e da Universidade Tecnológica Federal do Paraná, o trabalho apresentado foi considerado \_\_\_\_\_ (aprovado/reprovado) pela banca examinadora. No caso de aprovação, a mesma está condicionada ao cumprimento integral das exigências da banca examinadora, registradas no verso desta ata, da entrega da versão final da dissertação em conformidade com as normas da UTFPR e da entrega da documentação necessária à elaboração do diploma, em até \_\_\_\_\_ dias desta data.

Ciente (assinatura do aluno): \_\_\_\_\_

(para uso da coordenação)

A Coordenação do PPGCA/UTFPR declara que foram cumpridos todos os requisitos exigidos pelo programa para a obtenção do título de Mestre.

Curitiba PR, \_\_\_\_/\_\_\_\_/\_\_\_\_

**"A Ata de Defesa original está arquivada na Secretaria do PPGCA".**

*Dedico este trabalho à minha família. Esposa e filho, irmãos, cunhadas, sobrinhas e sobrinhos, e meus pais “seu” João e “dona” Rosa, que apesar do pouco estudo me proporcionaram valorosos ensinamentos, me ensinando a ser gente, do lado do bem.*

*Dedico também à minhas “amigas de estimação” Nala, Lulu e em especial a Laika, que após quase 15 anos de convivência, me deixou, dias após a defesa deste trabalho.*

*Saudade minha fiel amiga.*



## AGRADECIMENTOS

Agradeço a todos que de alguma forma colaboraram para que fosse possível o desenvolvimento deste trabalho.

A toda minha família, em especial, aos meus pais João e Rosa e irmãos pelo apoio que me dão desde sempre.

À minha esposa Viviane por compreender minhas intermináveis horas em frente ao computador, e pela dedicação ao meu filho Bruno, que é a coisa mais preciosa da minha vida, capaz de com apenas um sorriso acalmar os meus dias mais complicados.

À minha orientadora Prof.<sup>a</sup> Dr.<sup>a</sup> Tania Mezzadri Centeno pela orientação e confiança depositada ainda no processo de entrevista para aceitação no mestrado.

Ao Prof. Dr. João Alberto Fabro e ao Prof. Dr. Ricardo Dutra da Silva pelas valiosas recomendações fornecidas durante os seminários de qualificação.

Ao Prof. Ph.D. Oge Marques Filho por gentilmente fornecer a carta de recomendação necessária para minha inscrição no programa de mestrado.

Ao Programa de Pós-Graduação em Computação Aplicada (PPGCA) da Universidade Tecnológica Federal do Paraná e seus professores e funcionários pela estrutura e apoio oferecidos.

Aos senhores A. Young, Blackmore, Bowie, Coverdale, Dylan, Dio, Dickinson, Gillan, Gilmour, Harrison, Iommi, J. Ramone, Jagger, Kilmister, M. Young, McCafferty, Morrison, Mercury, Osbourne, Page, Plant, Ramalho, Richards, Seixas, Simmons, Stanley, Vecchione, Townshend, Waters, entre outros formidáveis músicos, e à rádio *online* KissFM por proporcionar que as obras destes músicos me acompanhassem, mantendo-me acordado, nas incontáveis madrugadas dedicadas à codificação e depuração dos algoritmos experimentados neste trabalho.

E acima de tudo a Deus, por me proporcionar saúde e me oferecer a possibilidade de perseguir meus objetivos.

*“Se A é o sucesso, então A é igual a X mais Y mais Z.  
O trabalho é X; Y é o lazer; e Z é manter a boca fechada.”*

Albert Einstein

*“Quando fizeres algo nobre e belo e ninguém notar, não fique triste.  
Pois o sol toda manhã faz um lindo espetáculo  
e no entanto, a maioria da platéia ainda dorme.”*

John Lennon

## RESUMO

GONÇALVES, Julio Cesar. **RECONHECIMENTO DE DÍGITOS EM IMAGENS DE MEDIDORES DE CONSUMO DE GÁS NATURAL UTILIZANDO TÉCNICAS DE VISÃO COMPUTACIONAL**. 102 f. Dissertação - Programa de Pós-Graduação em Computação Aplicada - Universidade Tecnológica Federal do Paraná. Curitiba, 2016.

Este trabalho propõe uma abordagem que emprega técnicas de processamento de imagens e classificação de padrões para o reconhecimento de dígitos apresentados no contador de consumo de medidores de gás natural. Tais imagens são obtidas em campo a partir de condições reais de operação, diferentemente da maioria das abordagens encontradas na literatura que se baseiam em imagens adquiridas em ambientes controlados. Inicialmente o contador de consumo é segmentado por técnicas de processamento de imagens. A segmentação é realizada com base no espaço de cor HSL da imagem, diferentemente da maioria dos trabalhos apresentados na literatura que utilizam imagens em tons de cinza. Em seguida os dígitos são individualmente segmentados e suas características extraídas de forma a compor uma base de conhecimento. Esta base serve de apoio para realizar a classificação e reconhecimento dos dígitos. Por fim, é feita uma comparação entre o desempenho dos classificadores KNN, SVM e ELM no reconhecimento de dígitos segmentados. Os resultados demonstram que as redes neurais ELM possuem um desempenho superior aos outros classificadores testados. Além disso, a metodologia mostrou-se promissora neste cenário, chegando a alcançar 95% de taxa de acerto no reconhecimento dos dígitos. Apresentando menos de 5% de falha no processo de segmentação do contador de consumo, considerando-se uma base com 903 imagens de medidores de gás. Diante da escassez de bases de imagens compatíveis com a finalidade desta pesquisa, tem-se como outro objetivo deste trabalho a disponibilização de uma base de dados contendo imagens de medidores de consumo de gás natural. Esta base é composta por imagens de medidores com tamanhos variados, obtidas em campo a partir de condições reais de operação. Fazem parte também desta base, imagens resultantes do processo de segmentação individual dos dígitos, com o objetivo de atender pesquisadores que pretendam apenas aplicar novos métodos de classificação.

**Palavras-chave:** Reconhecimento Ótico de Caracteres (OCR). Redes Neurais Artificiais. Máquina de Aprendizado Extremo (ELM). Máquinas de Vetores de Suporte (SVM).



## ABSTRACT

GONÇALVES, Julio Cesar. **DIGIT RECOGNITION IN IMAGES OF NATURAL GAS CONSUMPTION METERS USING COMPUTER VISION TECHNIQUES**. 102 f. Dissertação - Programa de Pós-Graduação em Computação Aplicada - Universidade Tecnológica Federal do Paraná. Curitiba, 2016.

This work proposes an approach that employs image processing techniques and pattern classification for the recognition of digits displayed in the counter of consumption of natural gas meters. These images are obtained in the field from actual operation conditions, unlike most of approaches in the literature that are based on images acquired in controlled environments. Initially, the counter of consumption is segmented by image processing techniques. The segmentation is performed based on the HSL color space of the image, unlike most of the works presented in the literature that uses images in grayscale. Then the digits are individually segmented and their features extracted in order to compose a knowledge base. This base serves as support to perform the classification and recognition of digits. Finally, a comparison is made between the performance of classifiers KNN, SVM and ELM in the recognition of segmented digits. The results demonstrate that the ELM neural networks have a performance superior to the other tested classifiers. Furthermore, the methodology showed to be promising in this scenario, reaching 95% success rate in recognition of digits. Presenting less than 5% fault in the process of segmentation the counter of consumption, considering a database with 903 images of gas meters. Given the scarcity of images compatible with the finality of this research, has as another objective of this work the availability of a database containing images of natural gas consumption meters. This base consists of images of gas meters with varying sizes, obtained from actual operating conditions. Also part of this base, resulting images of the individual segmented digits process, in order to meet researchers who wish to apply new methods of classification.

**Keywords:** Optical Characters Recognition (OCR). Artificial Neural Networks. Extreme Learning Machine (ELM). Support Vector Machine (SVM).

## LISTA DE FIGURAS

Figura 1. Representação da imagem em uma matriz bidimensional. ....	7
Figura 2. Representação da escala de níveis de cinza. ....	7
Figura 3. Tipos de vizinhança. ....	9
Figura 4. Modelos de cores.....	10
Figura 5. Representação do modelo de cores RGB. ....	10
Figura 6. Representação do modelo de cores HSL. ....	11
Figura 7. Etapas de um sistema de processamento de imagens. ....	13
Figura 8. Equalização de histograma.....	16
Figura 9. Aplicação do filtro da mediana por uma máscara 3x3. ....	17
Figura 10. Aplicação do filtro Gaussiano ....	18
Figura 11. Princípio da transformada de Hough ....	19
Figura 12. Transformada de Hough.....	20
Figura 13. Exemplo de histograma. ....	22
Figura 14. Aplicação de operadores para detecção de bordas na imagem.....	24
Figura 15. Rotulagem de componentes conexos.....	28
Figura 16. Interpolação bilinear.....	30
Figura 17. Esquema representativo do vetor de comportamento. ....	32
Figura 18. Exemplo de classificação via método SVM. ....	35
Figura 19. Estrutura de um neurônio artificial. ....	36
Figura 20. Estrutura de uma RNA de múltiplas camadas. ....	37
Figura 21. Arquitetura da rede neural ELM.....	39
Figura 22. Imagens de medidores em situações do cotidiano.....	64
Figura 23. Área vermelha apresentada no contador de consumo dos medidores ....	65
Figura 24. Resumo dos passos e principais técnicas utilizadas no método. ....	66
Figura 25. Passos para segmentação da ROI. ....	67
Figura 26. Limiarização de <i>pixels</i> vermelhos e <i>pixels</i> pretos.....	68
Figura 27. Aplicação do filtro de suavização em elemento do vetor de projeções ....	69
Figura 28. Projeções horizontais de pixels. (a) Original; (b) Suavizada. ....	69
Figura 29. Forma de localização de picos e vales no vetor de projeções. ....	70
Figura 30. Identificação de picos e vales na projeção suavizada. ....	70
Figura 31. Segmentação do resultado da limiarização na altura ....	71

Figura 32. Delimitação da largura da ROI.....	72
Figura 33. Ajuste de inclinação da ROI.....	73
Figura 34. Segmentação da altura definitiva da ROI .....	73
Figura 35. Passos para segmentação dos dígitos. ....	74
Figura 36. Filtro <i>top-hat</i> .....	75
Figura 37. Eliminação de ruídos .....	76
Figura 38. Segmentação dos dígitos .....	76
Figura 39. Ajuste pelo rótulo de dimensão máxima .....	77
Figura 40. Resumos dos passos para o reconhecimento dos dígitos. ....	78
Figura 41. Tela da aplicação que atribui rótulos aos dígitos segmentados. ....	79
Figura 42. Exemplos de falha na localização da ROI. ....	84
Figura 43. Amostras de imagens do grupo utilizado para validar o método. ....	86
Figura 44. Identificação de leitura. ....	87
Figura 45. Matriz de confusão.....	87
Figura 46 - Ciclômetro que realiza a contagem do volume de gás.....	90

## LISTA DE TABELAS

Tabela 1. Máscara de operadores para detecção de bordas. ....	24
Tabela 2. Classes de distribuição do vetor de comportamento. ....	32
Tabela 3. Mapeamento de características extraídas dos dígitos.....	80
Tabela 4. Taxa de acertos na localização da ROI. ....	83
Tabela 5. Taxas de reconhecimento obtidas com o experimento.....	85
Tabela 6. Percentuais de diagnóstico de leitura e reconhecimento de dígitos.....	87
Tabela 7. Tempo médio de execução das fases do método. ....	88
Tabela 8. Taxas de reconhecimento obtidas nos trabalhos relacionados. ....	88

## LISTA DE QUADROS

Quadro 1. Passos para rotulagem de componentes conexos. ....	28
Quadro 2. Resumo de técnicas relatadas na literatura. ....	56
Quadro 3. Contabilização das técnicas relatadas na literatura. ....	60

## LISTA DE ABREVIATURAS, SIGLAS E ACRÔNIMOS

ACC	Análise de Componentes Conexos
Arte-Lab	<i>Applied Recognition Technology Laboratory</i>
BDR	Base de Dígitos Rotulados
CCD	<i>Charge-coupled Device</i>
CMOS	<i>Complementary Metal Oxide Semiconductor</i>
COMPAGAS	Companhia Paranaense de Gás
CPU	<i>Central Processing Unit</i>
DSP	<i>Digital Signal Processor</i>
BP	<i>Back-propagation</i>
EE	Elemento Estruturante
ELM	<i>Extreme Learning Machine</i>
FPGA	<i>Field Programmable Gate Array</i>
FSM	<i>Finite State Machines</i>
GB	<i>Gigabyte</i>
GHz	<i>Gigahertz</i>
HSL	<i>Hue, Saturation and Lightness</i>
HSV	<i>Hue, Saturation and Value</i>
KNN	<i>K-Nearest Neighbours</i>
MLP	<i>Multilayer Perceptron</i>
NCC	<i>Normalized Cross-Correlation</i>
OCR	<i>Optical Character Recognition</i>
PC	<i>Personal Computer</i>
pixel	<i>picture element</i>
PPGCA	Programa de Pós-Graduação em Computação Aplicada
RAM	<i>Random-access memory</i>
RGB	<i>Red, Green and Blue</i>
RNA	Rede Neural Artificial
ROI	<i>Region of Interest</i>
SLFN	<i>Single-hidden Layer Feedforward Network</i>
SVM	<i>Support Vector Machine</i>

## SUMÁRIO

1 INTRODUÇÃO .....	1
1.1 JUSTIFICATIVA .....	2
1.2 OBJETIVOS .....	3
1.2.1 Objetivos gerais .....	4
1.2.2 Objetivos específicos .....	4
1.2.3 Contribuições esperadas .....	4
1.3 ESTRUTURA DO TRABALHO .....	5
2 FUNDAMENTAÇÃO TEÓRICA .....	6
2.1 INTRODUÇÃO .....	6
2.2 REPRESENTAÇÃO DE IMAGENS DIGITAIS .....	6
2.2.1 Histograma .....	7
2.2.2 Vizinhaça e Conectividade .....	8
2.3 MODELOS DE REPRESENTAÇÃO DE CORES .....	9
2.3.1 Modelo RGB .....	10
2.3.2 Modelo HSL .....	11
2.3.3 Conversão entre Modelo RGB e HSL .....	12
2.4 ETAPAS DE UM SISTEMA DE PROCESSAMENTO DIGITAL DE IMAGENS ...	13
2.5 PRÉ-PROCESSAMENTO .....	14
2.5.1 Transformação em Tons de Cinza .....	15
2.5.2 Equalização de Histograma .....	15
2.5.3 Suavização .....	16
2.5.4 Transformada de <i>Hough</i> .....	18
2.6 SEGMENTAÇÃO .....	21
2.6.1 Limiarização .....	21
2.6.2 Detecção de Bordas .....	23

2.6.3	Projeção Horizontal e Vertical.....	24
2.6.4	Morfologia Matemática Binária.....	25
2.6.4.1	Operador de Erosão .....	26
2.6.4.2	Operador de Dilatação .....	26
2.6.4.3	Operador de Abertura .....	26
2.6.4.4	Operador de Fechamento .....	27
2.6.4.5	Operadores <i>Top-hat</i> e <i>Bottom-hat</i> .....	27
2.6.4.6	Operador de Gradiente Morfológico.....	27
2.6.5	Análise de Componentes Conexos.....	28
2.6.6	Interpolação Bilinear .....	29
2.7	REPRESENTAÇÃO E DESCRIÇÃO .....	30
2.7.1	Características Estruturais via Projeções .....	30
2.7.2	Zoning .....	31
2.7.3	Vetor de Comportamento.....	31
2.8	TÉCNICAS PARA RECONHECIMENTO DE PADRÕES E CLASSIFICAÇÃO ...	33
2.8.1	K-vizinhos mais Próximos .....	33
2.8.2	<i>Support Vector Machine</i> .....	34
2.8.3	Redes Neurais Artificiais .....	35
2.8.4	<i>Extreme Learning Machine</i> .....	38
2.8.5	Avaliação de Métodos de Classificação.....	41
2.8.5.1	Método de Validação <i>Holdout</i> .....	41
2.8.5.2	Método de Validação Cruzada <i>K-fold</i> .....	41
2.8.5.3	Método de Validação <i>Leave-one-out</i> .....	42
2.9	CONSIDERAÇÕES FINAIS .....	42
3	ESTADO DA ARTE.....	43
3.1	INTRODUÇÃO .....	43
3.2	RECONHECIMENTO DE DÍGITOS APRESENTADOS EM MEDIDORES .....	43



3.3 CONSIDERAÇÕES FINAIS .....	55
4 MATERIAIS E MÉTODOS .....	63
4.1 INTRODUÇÃO .....	63
4.2 MATERIAIS.....	63
4.3 MÉTODOS.....	65
4.3.1 Segmentação da área contendo o contador de consumo .....	66
4.3.1.1 Conversão do espaço de cores RGB para HSL .....	67
4.3.1.2 Análise da projeção horizontal e vertical dos <i>pixels</i> .....	68
4.3.1.3 Análise de picos e vales.....	69
4.3.1.4 Segmentação da região de interesse na largura .....	71
4.3.1.5 Ajuste da inclinação .....	72
4.3.1.6 Segmentação da região de interesse na altura .....	73
4.3.2 Segmentação dos dígitos.....	74
4.3.6 Reconhecimento dos dígitos.....	77
4.3.6.1 Construção da base de amostras de dígitos.....	78
4.3.6.2 Extração de atributos .....	79
4.3.6.3 Classificação dos dígitos.....	80
4.4 CONSIDERAÇÕES FINAIS .....	81
5 RESULTADOS E DISCUSSÃO .....	83
5.1 INTRODUÇÃO .....	83
5.2 RESULTADOS OBTIDOS.....	83
5.5 DISCUSSÕES GERAIS .....	88
6 CONCLUSÕES E PERSPECTIVAS PARA TRABALHOS FUTUROS .....	92
6.1 CONCLUSÕES E DISCUSSÕES .....	92
6.2 TRABALHOS FUTUROS .....	95
REFERÊNCIAS .....	97

## 1 INTRODUÇÃO

O crescimento avançado da tecnologia digital, em conjunto com o desenvolvimento de novos algoritmos, tem permitido um número cada vez maior de aplicações que envolvem processamento digital de imagens. Como exemplo de áreas em que sistemas de processamento digital de imagens são aplicados podemos citar as áreas de medicina, biologia, sensoriamento remoto, astronomia, microscopia, arqueologia, segurança e vigilância, área militar e de automação industrial (PEDRINI e SCHWARTZ, 2008).

Nos dias atuais, o uso de dispositivos de captura de imagens, dentre os quais podem ser incluídos modernos e multifuncionais dispositivos móveis como celulares e tablets, têm se tornado cada vez mais comum. O interesse na extração de informações textuais de imagens do cotidiano capturadas por estes dispositivos tem aumentado (OBINATA e DUTTA, 2007; JUNG; KIM; JAIN, 2004).

Informações textuais apresentadas em imagens constituem um importante aspecto para contribuir com o entendimento destas imagens. A extração de texto em imagens capturadas em seu cenário natural apresenta muitas questões desafiadoras. Vários fatores influenciam a captura da imagem como: o ângulo de captura, iluminação excessiva, reflexos, bem como características do próprio texto apresentado como o tamanho da fonte, cor e alinhamento (ZHANG et al., 2002).

Sistemas de processamento digital de imagens têm um campo vasto em aplicações que analisam imagens no ambiente cotidiano. Extrair informações de imagens contendo um medidor de consumo de gás natural enquadra-se neste contexto.

Uma das fases do processo de faturamento em empresas distribuidoras de gás natural é a coleta do volume consumido pelos clientes. Aplicações voltadas para execução desta tarefa se tornam relevantes, uma vez que podem contribuir com a melhoria deste processo em distribuidoras que ainda recorrem à utilização de anotação manuscrita para coletar o valor consumido mensalmente pelos clientes.

É sabido que existem atualmente no mercado dispositivos medidores considerados “inteligentes”, que possuem recursos para enviar as informações de coleta de consumo por telemetria ou mesmo sistemas de cartões pré-pagos. Porém, segundo (HAMERSCHMIDT, 2012), o modelo atual que ainda se baseia em leituras manuais com mão-de-obra de leiturista, é economicamente eficiente para o sistema de distribuição de energia vigente. Tornando este tipo de mão de obra ainda considerado, em concessionárias que não consigam arcar com o custo da substituição ou adaptação de um parque de milhares de medidores.

Por outro lado, persistir neste modo manuscrito de operação leva à necessidade da transcrição de forma manual de milhares de dados de coleta para o sistema de faturamento. Esse tipo de abordagem pode gerar, entre outros problemas, atrasos e incorreções no faturamento, e até mesmo fraudes durante o transcorrer do processo.

A identificação deste problema motivou esta pesquisa, que teve como objeto de estudo a extração e reconhecimento de textos apresentados em equipamentos medidores de consumo de gás natural.

## 1.1 JUSTIFICATIVA

Apesar de se apresentar como uma atividade característica de reconhecimento ótico de caracteres (OCR) tradicional, a identificação de texto em objetos, como no caso de medidores, pode apresentar peculiaridades decorrentes de ruídos adquiridos na aquisição da imagem. Estes ruídos podem ser resultantes de diversas fontes, como a falta ou excesso de iluminação do ambiente, tipo e resolução do sensor da câmera, condição climática no momento da captura, sombras, oclusões, posicionamento da câmera em relação ao objeto capturado, entre outras interferências.

A apresentação destas peculiaridades pode atrapalhar a interpretação e reconhecimento de informações dispostas na imagem, inviabilizando a utilização de soluções tradicionais de sistemas de OCR, como pode ser observado no trabalho de (NODARI; GALLO, 2011), que utilizou uma biblioteca padrão de OCR denominada *Tesseract* (SMITH, 2007) obtendo uma taxa de 58% de acerto para o

reconhecimento dos dígitos. Este resultado, segundo o autor, ressaltou a complexidade da tarefa de OCR neste domínio de aplicação.

Em vista disto, do ponto de vista científico, este trabalho se justifica pelo propósito de encontrar um conjunto de algoritmos baseado em técnicas de processamento de imagens e reconhecimento de padrões, que seja apropriado para execução da tarefa de reconhecimento de dígitos apresentados em imagens de medidores de gás natural, colaborando assim para o avanço do estado-da-arte de detecção e reconhecimento de dígitos.

Do ponto de vista da indústria, justifica-se pelo esforço no sentido de proporcionar uma melhoria no processo de coleta de dados, bem como oferecer uma alternativa ao custo de substituição ou adaptação do parque de medidores existente nas distribuidoras.

## 1.2 OBJETIVOS

O propósito deste trabalho foi o de buscar abordagens robustas e de baixo custo computacional, capazes de serem reproduzidas em aplicações voltadas para dispositivos móveis, permitindo que a tarefa de reconhecimento de dígitos contidos no contador de volume de gás seja executada de forma eficaz.

Para tanto, foram empregadas diversas técnicas voltadas ao processamento digital de imagens com o objetivo de segmentar os dígitos contidos no contador de consumo. Para o método de classificação dos dígitos foi empregada a técnica denominada de Máquina de Aprendizagem Extrema (*Extreme Learn Machine – ELM*), um método desenvolvido por (HUANG; ZHU; SIEW, 2004) que apresenta como uma de suas principais características a velocidade de treinamento comparado aos métodos tradicionais de classificação.

Além de utilizar de forma prática ELM para o reconhecimento de dígitos, este trabalho apresenta também um estudo comparativo entre ELM e dois métodos de classificação clássicos encontrados na literatura (*K-Nearest Neighbours - KNN* e *Support Vector Machine – SVM*). Este estudo comparativo busca demonstrar a

eficiência prometida pelo método ELM relacionada ao baixo custo computacional para o reconhecimento dos dígitos, em comparação aos métodos KNN e SVM.

### 1.2.1 Objetivos gerais

O objetivo geral desta pesquisa consiste em obter um conjunto de algoritmos adequado para a segmentação de regiões de interesse e reconhecimento de padrões, aplicados ao domínio do problema da coleta de dados de consumo em medidores de energia.

### 1.2.2 Objetivos específicos

- Aquisição de massa de dados composta de imagens para treino e teste de algoritmos de processamento de imagens e reconhecimento de padrões.
- Obtenção de algoritmos para pré-processamento e segmentação de imagens, levando-se em conta o espaço de cores da imagem para localização da região de interesse, representada pelo contador de consumo de gás natural.
- Obtenção de algoritmos de extração de atributos e reconhecimento de padrões para interpretação do valor consumido destacado no contador de consumo de gás natural.

### 1.2.3 Contribuições esperadas

- Contribuir com a construção de um banco de imagens (treino e teste) que fique disponível para que outros pesquisadores elaborem futuras comparações com novas abordagens de algoritmos de processamento de imagem e reconhecimento de padrão.
- Contribuir com o desenvolvimento de um algoritmo que efetue a segmentação baseada no espaço de cores da imagem, diferentemente da maioria das abordagens da literatura que partem de imagens em tons de cinza para segmentação. Além disso, obter o reconhecimento dos dígitos utilizando um

algoritmo de reconhecimento de padrões, colaborando assim para o avanço do estado-da-arte de detecção e reconhecimento de dígitos.

- Apresentar um estudo comparativo entre o método de classificação ELM e os métodos clássicos de classificação KNN e SVM, métodos que já foram abordados em outros trabalhos envolvendo o reconhecimento de dígitos apresentados em medidores de energia.

### 1.3 ESTRUTURA DO TRABALHO

Este documento está organizado em cinco seções. Nesta seção, é apresentada a contextualização do problema que se pretende resolver utilizando técnicas de processamento de imagens e reconhecimento de padrões, bem como a justificativa do trabalho, objetivos e contribuições esperadas.

A seção 2 apresenta uma fundamentação teórica sobre técnicas utilizadas em processamento digital de imagens e reconhecimento e interpretação de padrões.

Na seção 3 são abordados os estudos publicados na literatura relacionados ao problema do reconhecimento de dígitos em instrumentos de medição. Nesta mesma seção é apresentada uma sumarização das técnicas abordadas em cada estudo.

Na seção 4 é apresentada a metodologia utilizada para o alcance dos objetivos.

Na seção 5 são apresentados os experimentos e resultados alcançados com base na metodologia proposta na seção anterior.

Concluindo, a seção 6 apresenta as considerações finais e sugestões de trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 INTRODUÇÃO

Sistemas de processamento digital de imagens consistem em um conjunto de técnicas que permitem a captura de imagens por equipamentos variados, como tomógrafos médicos, câmeras fotográficas, satélites e outros, bem como a representação e transformação destas imagens. O propósito é extrair e identificar informações contidas nas imagens de forma automática por meio de máquinas (PEDRINI e SCHWARTZ, 2008). Nas próximas seções descrevemos alguns conceitos e técnicas empregadas em processamento digital de imagens, bem como técnicas voltadas para o reconhecimento e interpretação de padrões, úteis na composição de sistemas desta natureza.

### 2.2 REPRESENTAÇÃO DE IMAGENS DIGITAIS

A representação de uma imagem é uma função  $f(x,y)$ , onde  $x$  e  $y$  são coordenadas espaciais. A função  $f(x,y)$  possui, comumente, um valor em níveis de cinza, ou cor, em qualquer ponto da imagem, o ponto de origem representado pela coordenada  $(0,0)$  fica situado no canto superior esquerdo da imagem (GONZALEZ E WOODS, 2002).

A Figura 1 ilustra a notação matricial para representação de uma imagem digital. Cada ponto desta matriz é denominado de *pixel* (do inglês *picture element*). O índice apontado pela letra  $m$  representa a linha na qual o *pixel* se encontra, enquanto que o índice apontado pela letra  $n$  aponta a coluna na qual está localizado o *pixel*. Supondo que a imagem contenha  $M$  linhas e  $N$  colunas,  $m$  terá seu valor variando de 0 até  $M-1$ , enquanto que  $n$  terá seu valor variando de 0 até  $N-1$ . Em imagens em tons de cinza de 8 *bits*, o *pixel* tem o seu valor variando de 0 (representando a cor preta) até 255 (representando a cor branca).

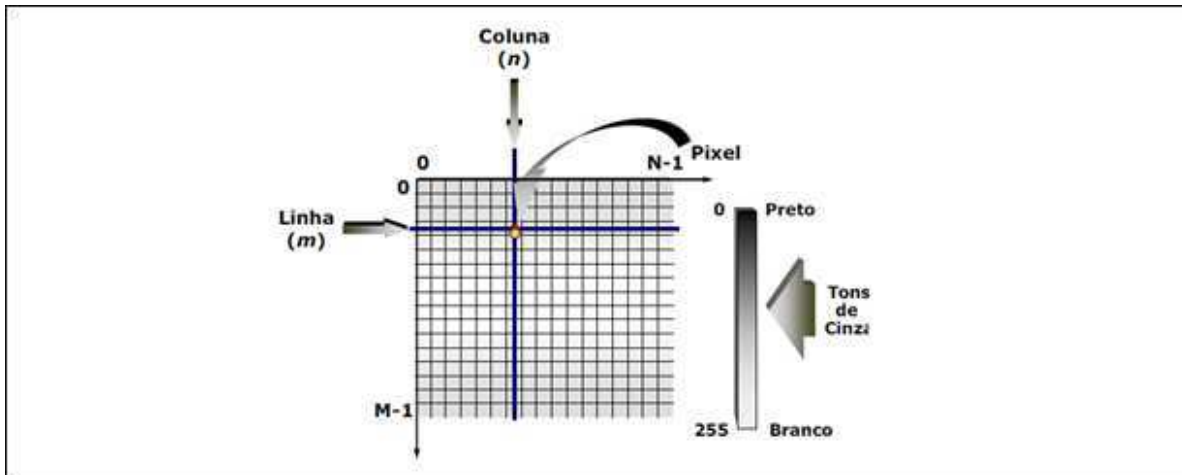


Figura 1. Representação da imagem em uma matriz bidimensional.

Fonte: De Queiroz e Gomes (2006).

### 2.2.1 Histograma

O histograma representa a distribuição dos níveis de cinza apresentados em uma imagem. A quantidade de cada nível de cinza é representada por um gráfico (ACHARYA; RAY, 2005). A Figura 2a exibe uma imagem em níveis de cinza de um medidor de consumo de gás, a Figura 2b apresenta o gráfico resultante do histograma correspondente à imagem do medidor. O gráfico fornece a informação sobre quantidade de *pixels* na imagem que correspondem a cada um dos valores possíveis da escala de níveis de cinza entre 0 e 255. Isto permite destacar as proporções da imagem que correspondem a cada um dos níveis de cinza.

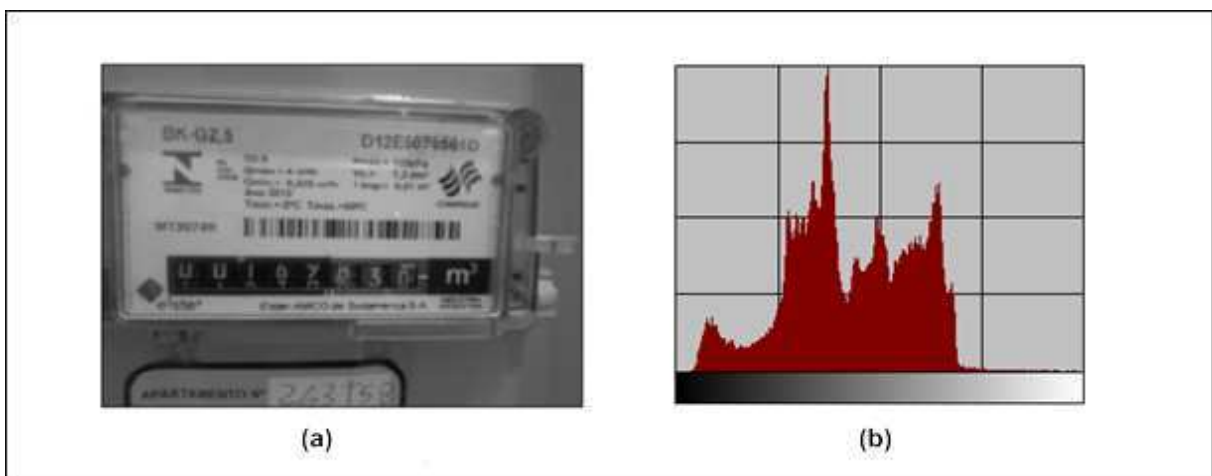


Figura 2. Representação da escala de níveis de cinza. (a) Imagem; (b) Histograma.



Por meio do histograma pode ser avaliado o contraste, bem como o brilho médio de uma imagem, indicando se esta é predominantemente clara ou escura. Escalas estreitas do nível de cinza indicam que a imagem é de baixo contraste, enquanto que um histograma com escala bem distribuída indica uma imagem com alto contraste (PEDRINI e SCHWARTZ, 2008).

Outras características da imagem podem também ser obtidas utilizando a análise do histograma, tais como o nível de cinza máximo, mínimo e médio, bem como a variância e o desvio padrão da distribuição dos níveis de cinza.

### 2.2.2 Vizinhaça e Conectividade

Existem vários relacionamentos importantes que podem ser observados entre os *pixels* de uma imagem digital. Estes dizem respeito à região pela qual determinado *pixel* é circundado, bem como a conexão existente entre este *pixel* e a região que o circunda.

Conforme descrito em Gonzalez e Woods (2002), um *pixel*  $p$  de coordenadas  $(x,y)$  possui dois *pixels* vizinhos horizontais e dois vizinhos verticais com coordenadas determinadas por  $(x-1, y)$ ,  $(x+1, y)$  e  $(x, y-1)$ ,  $(x, y+1)$  respectivamente e cuja formação é denominada vizinhaça-4. Há também uma vizinhaça diagonal composta das coordenadas  $(x-1, y-1)$ ,  $(x-1, y+1)$ ,  $(x+1, y-1)$  e  $(x+1, y+1)$ . A união destas vizinhaças forma o que se convencionou chamar de vizinhaça-8.

Deve-se observar que alguns dos *pixels* vizinhos não estarão presentes na imagem quando o *pixel*  $p$  estiver localizado na borda da imagem. Isto pode ser mais bem entendido analisando-se a Figura 3 que demonstra visualmente o conceito de vizinhaça de um *pixel*.

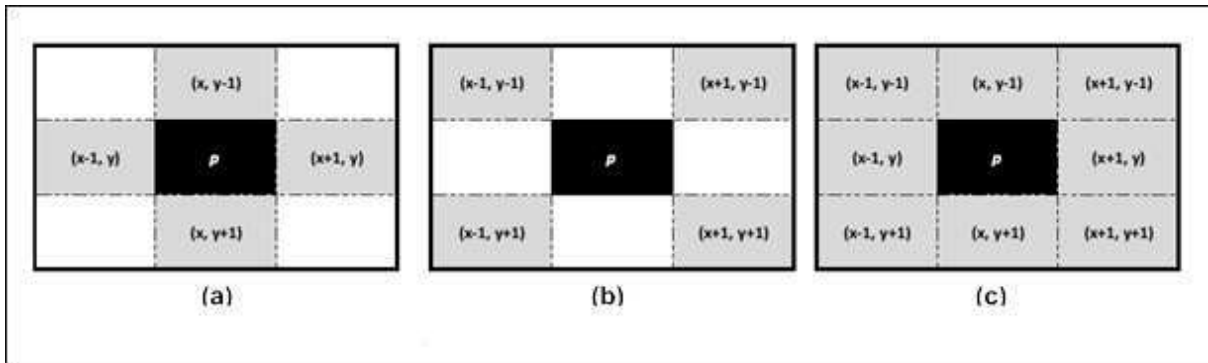


Figura 3. Tipos de vizinhança. (a) Vizinhança-4; (b) Vizinhança diagonal; (c) Vizinhança-8.

Ainda segundo os autores, a conectividade entre *pixels* é um conceito importante no processamento de imagens, em especial para o estabelecimento das bordas de objetos presentes em uma imagem. Para estabelecer se dois *pixels* são conexos, é preciso determinar se são vizinhos e se seus níveis de cinza são similares. Deve-se observar que, há a possibilidade de determinação da conexidade por meio da utilização da vizinhança-4, resultando em uma conectividade-4, ou utilizando a vizinhança-8, resultando em uma conectividade-8.

### 2.3 MODELOS DE REPRESENTAÇÃO DE CORES

Segundo Gonzalez e Woods (2002), a cor é um poderoso descritor de características que frequentemente simplifica o processo de identificar um objeto em determinada cena. O propósito de um modelo de cores é proporcionar uma forma padrão e de aceite geral que facilite a especificação de cores.

Modelos de cores podem ser classificados como aditivos ou subtrativos (PEDRINI e SCHWARTZ, 2008). No primeiro, a cor branca é gerada pela adição das cores primárias verde, vermelha e azul. A cor preta corresponde à ausência de cor. Já no modelo subtrativo, a cor preta é gerada pela combinação das cores primárias ciano, magenta e amarelo, e a cor branca corresponde à ausência destas cores. A Figura 4 ilustra a combinação de cores nos modelos aditivos e subtrativos.

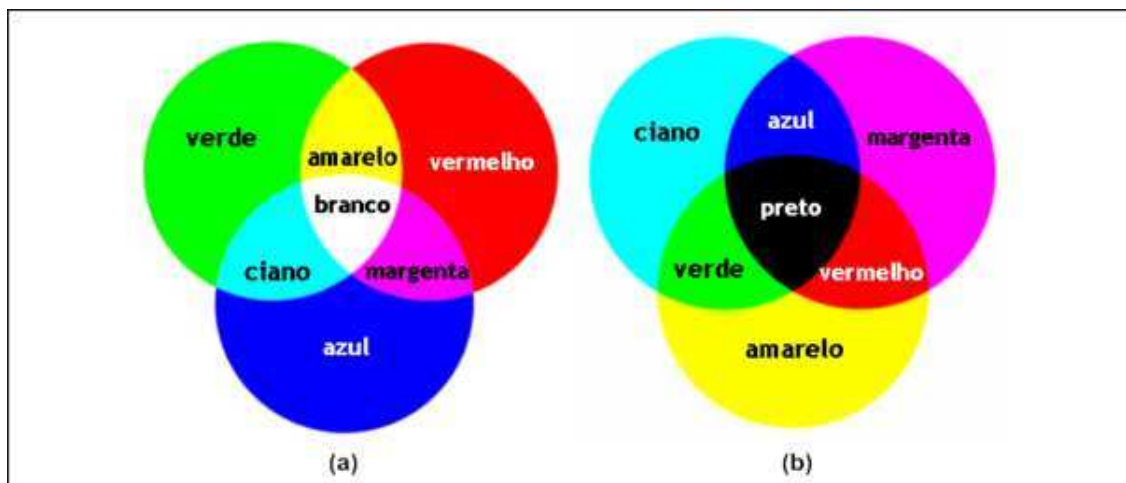


Figura 4. Modelos de cores. (a) Aditivo; (b) Subtrativo.

Fonte: Adaptado de Pedrini e Schwartz (2008).

Nas próximas seções são descritas as características básicas dos modelos tratados nesta pesquisa, que são os modelos aditivos de representação de cores RGB e HSL, bem como o método de conversão entre estes modelos.

### 2.3.1 Modelo RGB

No modelo RGB cada cor aparece nos seus componentes de espectro primários, vermelho (R, *red*), verde (G, *green*) e azul (B, *blue*). Neste modelo, a escala de cinza estende-se do preto até o branco (GONZALEZ e WOODS, 2002). A Figura 5 ilustra o subespaço de cores no modelo RGB.

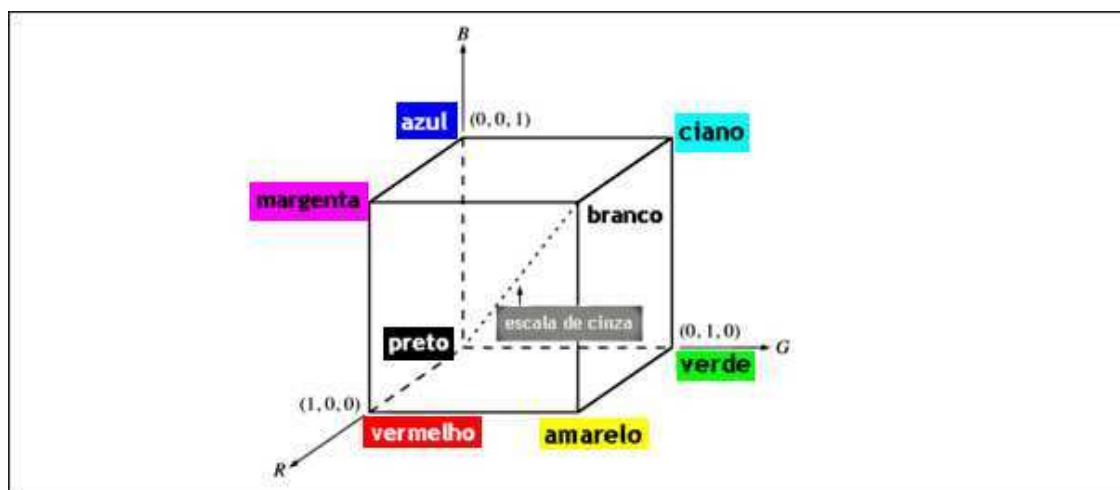


Figura 5. Representação do modelo de cores RGB.

Fonte: Adaptado de Gonzalez e Woods (2002).

Conforme pode ser observado na Figura 5, os valores RGB localizam-se nos três cantos dos eixos principais; ciano, magenta e amarelo estão localizados nos outros três cantos; preto está na origem e branco está no canto mais distante da origem.

### 2.3.2 Modelo HSL

O modelo HSL é definido pelos parâmetros matiz (H, *hue*), saturação (S, *saturation*) e luminosidade (L, *lightness*), e sua representação gráfica é formada por dois cones de altura 1 cujas bases são coincidentes (PEDRINI e SCHWARTZ, 2008), conforme ilustrado na Figura 6.

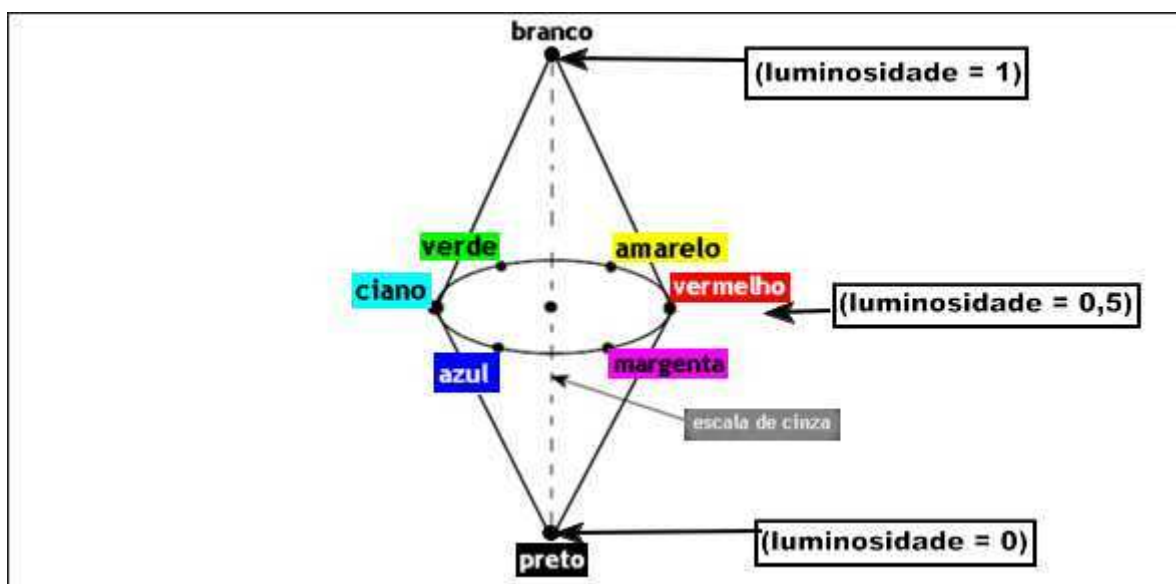


Figura 6. Representação do modelo de cores HSL.

Fonte: Pedrini e Schwartz (2008).

O matiz é determinado pelos pontos no círculo da borda das duas bases comuns aos cones. A saturação varia conforme a distância do eixo entre 0 e 1. A luminosidade varia de preto (valor 0) a branco (valor 1) ao longo do eixo comum aos dois cones, onde se encontra a escala de cinza.

O espaço HSL é semelhante ao espaço HSV abordado no trabalho de Jabba et al. (2012), pois levam o mesmo valor numérico para o matiz (H). O matiz determina a cor ou tonalidade (vermelho, marrom, amarelo, etc.)

No entanto, no modelo HSL trabalha-se com a luminância ou leveza (L) atribuindo altos valores de cores a esta quanto mais se aproximar da cor branca (IBRAHEEM et al., 2012). Enquanto que no modelo HSV trabalha-se com o brilho (V) que pode ser percebido como a "quantidade de luz", e determina a intensidade de cor.

A saturação é diferente, pois dependendo do modelo é redimensionada para caber no sentido de brilho (V) no qual uma cor pura é igual à luminosidade de um cinza médio, ou no sentido de leveza (L) no qual uma cor pura é igual à luminosidade do branco.

### 2.3.3 Conversão entre Modelo RGB e HSL

O modelo RGB, é o modelo de cores comumente resultante do processo de aquisição de imagens por modelos de câmeras populares. Segundo Pedrini e Schwartz (2008), por meio das equações (1), (2) e (3), podemos converter uma imagem capturada no modelo RGB para o modelo HSL.

$$H = \begin{cases} 0, & \text{se } M = m \\ 60 \frac{(G - B)}{(M - m)}, & \text{se } M = R \\ 60 \frac{(B - R)}{(M - m)} + 120, & \text{se } M = G \\ 60 \frac{(R - G)}{(M - m)} + 240, & \text{se } M = B \end{cases} \quad (1)$$

$$S = \begin{cases} 0, & \text{se } M = m \\ \frac{(M - m)}{(M + m)}, & \text{se } 0 < L \leq 0,5 \\ \frac{(M - m)}{2 - (M + m)}, & \text{se } L > 0,5 \end{cases} \quad (2)$$

$$L = \frac{(M + m)}{2} \quad (3)$$

Aonde,  $m=\min(R,G,B)$  e  $M=\max(R,G,B)$ . A luminância  $L$  e a saturação  $S$  estão normalizadas entre 0 e 1. O matiz  $H$  varia entre 0 e 360 graus.

## 2.4 ETAPAS DE UM SISTEMA DE PROCESSAMENTO DIGITAL DE IMAGENS

De acordo com Pedrini e Schwartz (2008), um sistema de processamento de imagens é composto por um conjunto de etapas, capaz de produzir determinado resultado a partir de um domínio do problema. A Figura 7 ilustra as etapas de um sistema de processamento de imagens.

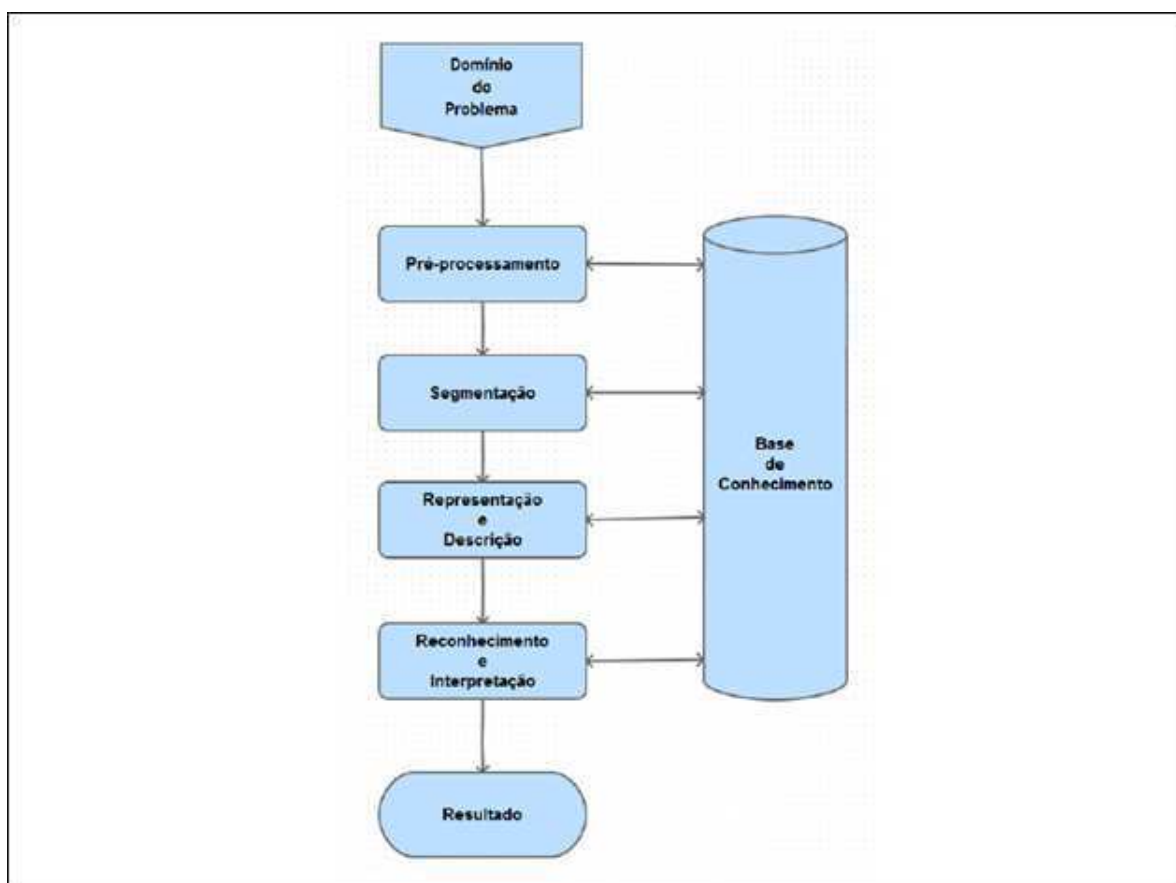


Figura 7. Etapas de um sistema de processamento de imagens.

Fonte: Adaptado de Pedrini e Schwartz (2008).

Na etapa de aquisição, obtém-se a imagem a ser processada por meio de dispositivos como, por exemplo, câmeras fotográficas ou câmeras de vídeo. Converte-se a imagem em uma representação adequada para ser utilizada nas próximas etapas, levando-se em conta aspectos como a escolha do tipo de sensor,

condições de iluminação, resolução e número de cores ou níveis de cinza da imagem.

Na etapa de pré-processamento, busca-se a melhoria da qualidade da imagem por meio da redução de possíveis ruídos provenientes da etapa de aquisição, empregando a suavização de propriedades da imagem, bem como a correção de aspectos como brilho e contraste.

Na etapa de segmentação realiza-se a identificação e extração da região ou das regiões de interesse (*ROI – Region of Interest*) apresentadas na imagem, geralmente baseando-se na detecção de descontinuidades que são identificadas como bordas ou regiões similares apresentadas na imagem.

Na etapa de representação e descrição, visa-se extrair determinados atributos ou características que possam discriminar classes de objetos que compõem a imagem.

A última etapa envolve o reconhecimento e a interpretação de componentes identificados em uma imagem. Nesta etapa é feita a classificação e seleção por meio de reconhecimento de padrões, baseado nas características apresentadas pelos descritores fornecidos pela etapa anterior.

Para execução das tarefas descritas em cada etapa é pressuposto um conhecimento prévio sobre o problema a ser resolvido, este conhecimento pode estar codificado na forma de uma base de conhecimento cuja complexidade e tamanho podem variar de forma significativa a depender da aplicação. O objetivo de utilizar a base de conhecimento é o de guiar o funcionamento de cada etapa de processamento a fim de executar determinada tarefa.

## 2.5 PRÉ-PROCESSAMENTO

Imagens reais com frequência apresentam degradações durante seu processo de aquisição e são comumente denominadas de ruídos (PEDRINI e SCHWARTZ, 2008). A função da etapa de pré-processamento em um sistema de processamento de imagens, é aprimorar a imagem que será utilizada nas etapas posteriores, buscando a eliminação da ocorrência destes ruídos.

As operações efetuadas para aprimorar uma imagem envolvem técnicas que visam melhorar a aparência ou converter a imagem para um formato mais adequado, permitindo a análise da mesma por uma máquina (PRATT, 2007). A seguir, são descritas algumas destas operações.

### 2.5.1 Transformação em Tons de Cinza

Imagens em tons de cinza com profundidade de 8 bits fornecem 256 níveis de cinza tendo a luminosidade representada por um único valor que varia do preto ao branco. Já para o caso de imagens coloridas, ao invés da intensidade de luz ser representada por um único valor, cada *pixel* é representado pela combinação de três valores no sistema RGB (vermelho, verde e azul), ou seja, três planos de 8 *bits* (24 *bits*), totalizando mais de 16 milhões de cores diferentes (GONZALEZ e WOODS, 2002).

Para a conversão de uma imagem colorida em uma imagem em escalas de cinza, a equação (4) pode ser utilizada, na qual  $R$  significa o valor de vermelho (*red*),  $G$  o valor de verde (*green*) e  $B$  o valor de azul (*blue*) que compõem o *pixel*.

$$EscalaCinza = 0,299R + 0,587G + 0,114B \quad (4)$$

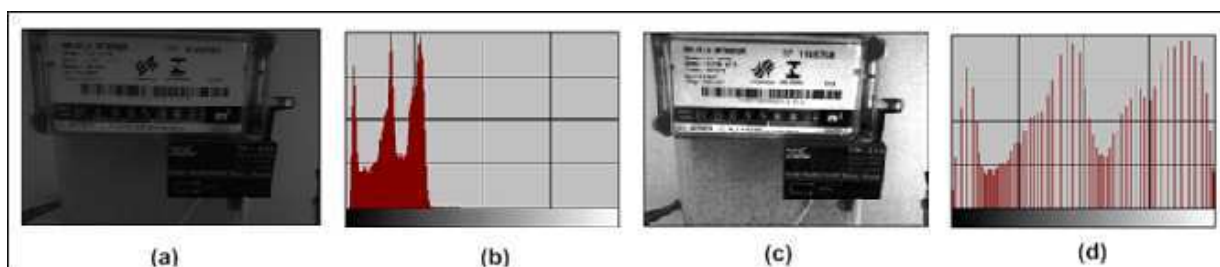
### 2.5.2 Equalização de Histograma

A operação de equalização de histograma modifica a imagem de tal forma que o resultado da transformação seja uma imagem com uma distribuição mais uniforme de níveis de cinza. A imagem apresentará os seus níveis aproximadamente com a mesma frequência (PEDRINI e SCHWARTZ, 2008). O histograma equalizado pode ser obtido pela equação (5), onde  $k$  é a intensidade no histograma equalizado,  $L$  é o valor máximo de intensidade na imagem,  $M$  e  $N$  são as dimensões da imagem e  $n_j$  é o número de *pixel* na imagem com valor de intensidade igual a  $j$ .

$$h_{eq(k)} = \frac{(L-1)}{MN} \sum_{j=0}^k n_j \quad (5)$$



A Figura 8 retrata o ajuste do contraste de uma imagem mediante a aplicação da operação de equalização de histograma.



**Figura 8. Equalização de histograma. (a) Imagem original; (b) Histograma da imagem original; (c) Imagem equalizada; (d) Histograma da imagem equalizada.**

### 2.5.3 Suavização

Operadores de suavização são utilizados para diminuir o efeito de ruídos originados do processo de aquisição da imagem. A operação para suavizar uma imagem é executada pela aplicação de filtros compostos de matrizes (máscaras) quadradas  $M \times M$ , que percorrem toda a imagem alterando o valor dos *pixels* de acordo com seus coeficientes. O cálculo do valor do *pixel* a ser alterado é baseado na média aritmética ponderada que corresponde à soma dos valores de entrada ao redor do *pixel*, cada qual multiplicado pelo coeficiente correspondente da máscara.

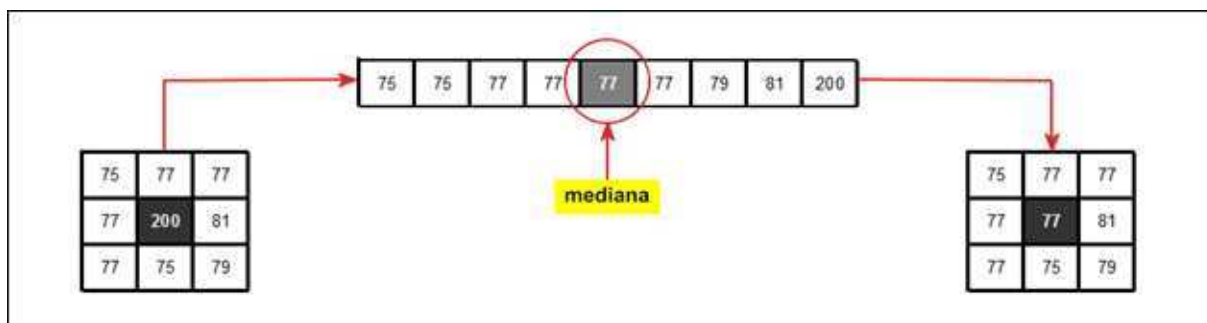
A filtragem linear de uma imagem  $f$  de tamanho  $M \times M$  por uma máscara  $w$  de tamanho  $m \times n$  ( $m$  e  $n$  usualmente ímpares) é dada pela expressão apresentada na equação (6), onde  $a=(m-1)/2$  linhas e  $b=(n-1)/2$  colunas (GONZALEZ e WOODS, 2002).

$$g(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x + s, y + t) \quad (6)$$

Filtros lineares como o filtro passa-baixa, são capazes de efetuar a suavização, no entanto, também podem causar o borramento e perda de detalhes da imagem como as bordas. Para contornar essa perda são utilizados filtros não lineares, pois estes não suavizam regiões próximas à borda de forma homogênea.

De acordo com Pedrini e Schwartz (2008), um dos filtros não lineares mais importantes é o filtro da mediana, o qual consiste em substituir a intensidade de

cada *pixel* pela mediana de intensidades da vizinhança do *pixel*. Isso é conseguido fazendo a ordenação crescente de intensidade dos *pixels* que estão sob a aplicação da máscara, substituindo o *pixel* central pelo valor mediano da ordenação. A Figura 9 ilustra a aplicação do filtro da mediana utilizando uma máscara de dimensão 3x3.



**Figura 9. Aplicação do filtro da mediana por uma máscara 3x3.**

Outro filtro que proporciona o borramento da imagem é o filtro Gaussiano. Gonzalez e Woods (2002) descrevem-no como uma das alternativas para suavizar imagens com o propósito de redução de ruídos. Segundo Pedrini e Schwartz (2008), os coeficientes da máscara utilizada na filtragem são derivados de uma função Gaussiana bidimensional. A equação (7) define esta função na qual o desvio padrão  $\sigma$  representa o grau de suavização.

$$g(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-(x^2 + y^2)}{2\sigma^2}\right) \quad (7)$$

Ainda segundo Pedrini e Schwartz (2008), o filtro Gaussiano possui diversas características úteis, com destaque para:

- Em duas dimensões, é um filtro de característica isotrópica, pois, devido ao filtro ser simétrico em relação à rotação, realizará a suavização com o mesmo grau em todas as direções.
- A suavização acontece por meio da substituição de cada *pixel* por uma média ponderada dos *pixels* vizinhos, onde o peso de um vizinho decresce com a distância do *pixel* central.

- O grau de suavização (largura) do filtro está relacionado com o valor do desvio padrão  $\sigma$ . Quanto maior  $\sigma$ , maior é a largura do filtro e maior é o seu grau de suavização.
- As funções Gaussianas são separáveis, permitindo que a convolução seja executada por meio da aplicação de um filtro Gaussiano unidirecional, seguido da aplicação do mesmo filtro orientado ortogonalmente, o que possibilita reduzir consideravelmente o número necessário de operações executadas.

A Figura 10 apresenta o resultado da aplicação de um filtro Gaussiano.



Figura 10. Aplicação do filtro Gaussiano. (a) Imagem original; (b) Imagem filtrada ( $\sigma = 7$ ).

#### 2.5.4 Transformada de *Hough*

Segundo Pedrini e Schwartz (2008), a área de análise de imagens apresenta com frequência problemas que consistem em detectar um conjunto de pontos que pertençam, por exemplo, a um segmento de reta.

No método proposto por Hough (1962) é possível detectar, em imagens binarizadas, formas passíveis de parametrização como linhas retas e circunferências. O método, denominado transformada de Hough, baseia-se na conversão de uma imagem binária em um espaço acumulador de parâmetros com  $n$ -dimensões, sendo  $n$  o número de parâmetros pelos quais o objeto é definido.

Na detecção de linhas, considera-se que para um ponto  $(x, y)$  pertencente a imagem existirá um conjunto de retas que interceptará este ponto e que são definidas pela equação (8).

$$y = mx + b \quad (8)$$

A equação anterior pode ser recombinaada tornando  $m$  e  $b$  parâmetros e transformando  $x$  e  $y$  em constantes. Tem-se então a equação (9) que é denominada como espaço de parâmetros ou espaço de Hough.

$$b = y - mx \quad (9)$$

Observando as equações anteriores, percebe-se que cada reta que passa pelo ponto  $(x, y)$ , irá corresponder a um ponto no espaço de Hough. De fato todos os pontos que são colineares no plano da imagem irão se interceptar em um mesmo ponto no espaço de parâmetros.

A Figura 11 retrata a relação de mapeamento para os pontos  $p1$ ,  $p2$  e  $p3$ . Dentre o conjunto de retas que passam por cada um destes pontos, existe apenas uma que passa simultaneamente pelos três pontos. Esta reta é então representada no espaço de Hough pelo ponto de intersecção  $(m, b)$ .

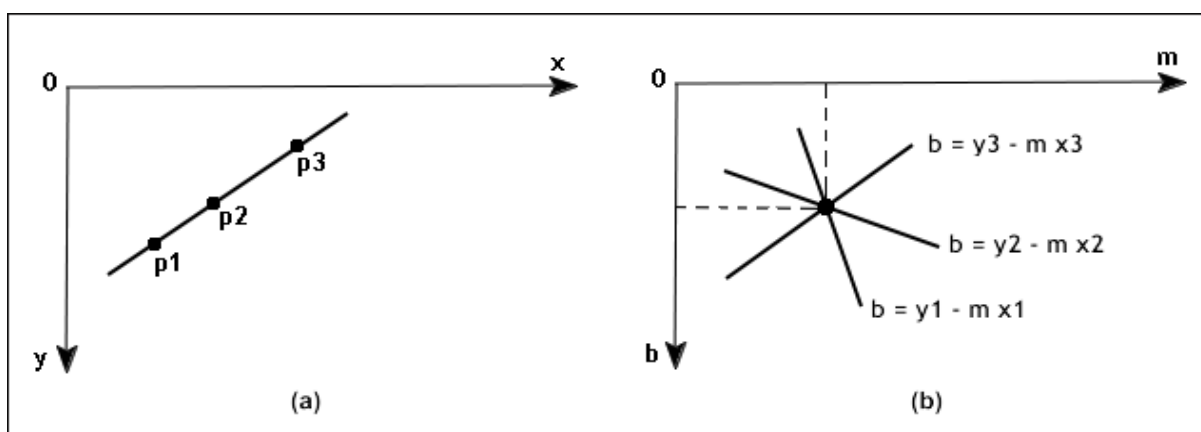


Figura 11. Princípio da transformada de Hough. (a) Plano da imagem; (b) Espaço de Hough.

A abordagem descrita anteriormente apresenta problemas quando as retas forem aproximadamente verticais com declividade tendendo ao infinito (PEDRINI E SCHWARTZ, 2008). A aplicação da equação (10) contornar este problema, pois define as retas na forma de coordenadas polares.

$$\rho = x \cos \theta + y \sin \theta \quad (10)$$

Onde:  $\rho$  é a distância perpendicular da origem (0,0) à reta, e  $\theta$  é o ângulo formado entre a reta perpendicular e o eixo x.

A definição das retas na forma de coordenadas polares implica na alteração do espaço de Hough, para o qual as retas passam a ser substituídas por curvas, porém sem alterar o funcionamento da técnica. A Figura 12 ilustra a representação da reta em forma polar, bem como exibe a nova representação do espaço de Hough.

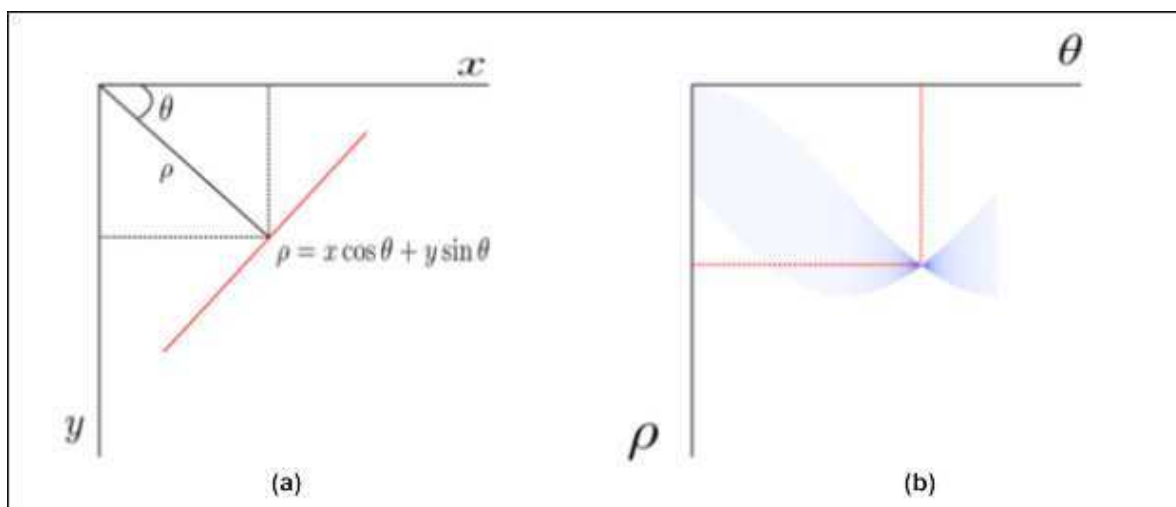


Figura 12. Transformada de Hough. (a) Reta na forma polar; (b) Espaço de Hough.

Como descrito no início desta seção, a detecção das retas deve ser feita a partir de uma imagem binarizada, cujos pontos de interesse podem ser determinados pela aplicação de um algoritmo para detecção de bordas, como por exemplo, o operador de Sobel. Em seguida utiliza-se um acumulador com base em uma matriz de pontos  $(\rho, \theta)$ , decorrendo um processo de votação no espaço de *Hough*. Para cada ponto de interesse é incrementada a respectiva entrada no acumulador e terminada esta análise são escolhidas as entradas de interesse com base num *threshold* previamente definido. O resultado deste processo irá determinar as retas existentes na imagem.

## 2.6 SEGMENTAÇÃO

Segundo Gonzalez e Woods (2002) o objetivo da segmentação é obter uma imagem subdividida em partes como forma de buscar o isolamento de conjuntos de *pixels* que representem as regiões de interesse (ROI) procuradas na imagem. Variadas técnicas podem ser empregadas para a tarefa de segmentação. Algumas destas técnicas são apresentadas nas próximas seções.

### 2.6.1 Limiarização

A segmentação por limiarização, também denominada de *Thresholding*, é um método para dividir a imagem resultando em duas regiões de interesse: objeto e fundo. O agrupamento dos *pixels* nestes dois grupos é feito por meio de um valor limiar de corte, cujos *pixels* com valores maiores que o limiar são classificados no grupo de objetos e os *pixels* restantes classificados no grupo de fundo. A equação (11) corresponde a forma tradicional de cálculo do limiar no qual  $R_1$  e  $R_2$  são os valores estipulados para os níveis de cinza da imagem binarizada, no caso utiliza-se 0 (para a cor preta) e 255 (para a cor branca), e  $T$  o valor limiar estimado. Uma das técnicas para estimativa do valor limiar  $T$  é descrita em Otsu (1979).

$$g(x, y) = \begin{cases} R_1 & \text{se } f(x, y) \leq T \\ R_2 & \text{se } f(x, y) > T \end{cases} \quad (11)$$

A limiarização pode ser feita por meio da inspeção do histograma da imagem. Para o caso de um histograma bi-modal a distribuição de fundo e objeto é bem definida. Já em casos em que o histograma não se apresenta bi-modal, é necessária a execução da limiarização por tentativa utilizando processos iterativos, na qual são testados diferentes limiares até produzir um resultado satisfatório de acordo com o observador.

A Figura 13(a) exhibe a representação de um histograma bi-modal e a Figura 13(b) um histograma que apresenta dois níveis de *limiar*.

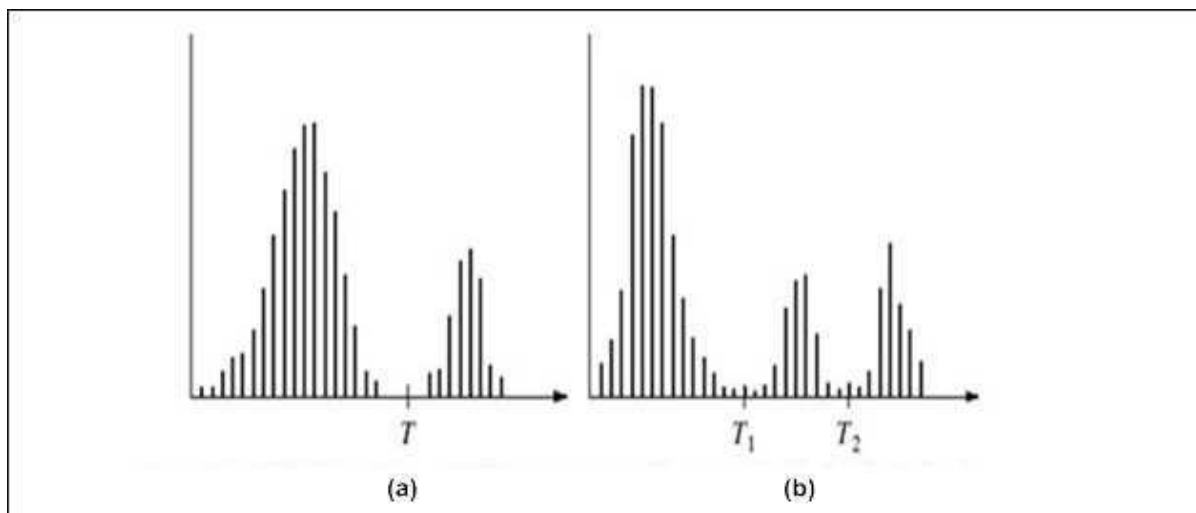


Figura 13. Exemplo de histograma. (a) Histograma bi-modal; (b) Histograma não bi-modal.

Como alternativa para o limiar de distribuições que não se apresentam na forma bi-modal encontra-se a limiarização adaptativa na qual o algoritmo tenta encontrar um valor de limiar realizando o cálculo utilizando pequenas máscaras de vizinhança (como por exemplo,  $3 \times 3$ ). Em cada janela é calculado um valor global para cada *pixel* central da máscara (limiar local). A principal vantagem do uso de técnicas, em relação ao limiar fixo, é a compensação de variações nas condições de iluminação da cena em que a imagem é gerada.

Um algoritmo apresentado na literatura para o cálculo da limiarização adaptativa é o algoritmo de Bernsen (1986), que calcula os valores de níveis de cinza presentes na vizinhança de cada *pixel* utilizando o seguinte procedimento.

Para cada *pixel* na posição  $(x,y)$ , o limiar  $T(x,y) = (P_{Menor} + P_{Maior})/2$  é usado, onde  $P_{Menor}$  e  $P_{Maior}$  correspondem ao mais baixo e mais alto valores de nível de cinza em uma vizinhança  $R \times R$  centrada em  $(x,y)$ . Para os casos em que a medida de contraste  $C(x,y) = (P_{Maior} - P_{Menor})$  for menor que um contraste mínimo  $L$ , então a vizinhança só consiste em uma das classes; preto ou branco.

Outra forma de realizar a limiarização adaptativa é utilizando o método de Niblack (1986). Este método determina o valor do limiar por meio do cálculo da média e do desvio padrão dos valores dos *pixels* presentes na janela de vizinhança, e a partir do cálculo destes valores aplica-se a equação (12), onde:

$T(x,y)$  é o valor limiar a ser determinado para o *pixel* em análise.

$\mu(x,y)$  é a média calculada dos *pixels* apresentados na janela.

$\sigma(x,y)$  é o desvio padrão calculado dos *pixels* apresentados na janela.

$k$  é um parâmetro que possibilita o controle da influência do desvio padrão das intensidades vizinhas sobre a decisão de limiar.

$$T(x,y) = \mu(x,y) + k\sigma(x,y) \quad (12)$$

Em resumo, percorre-se a imagem calculando, dentro da janela de vizinhança, a soma do resultado da média com o resultado do desvio padrão multiplicado por  $k$ . Assim, um valor para o limiar  $T$  é determinado e caso seja menor que o valor do *pixel* em análise, esse *pixel* recebe zero, senão recebe um.

### 2.6.2 Detecção de Bordas

Conforme Pedrini e Schwartz (2008), a definição de borda é dada como uma fronteira ou limite entre duas regiões na imagem que apresenta as propriedades de níveis de cinza relativamente distintas.

As bordas podem ser determinadas com base nestas discontinuidades de níveis de cinza. Existem variadas técnicas e abordagens empregadas para a tarefa de detecção de bordas. Alguns exemplos de operadores clássicos são: o operador de Roberts, o operador de Sobel e o operador de Prewitt.

As máscaras utilizadas para determinação das bordas de acordo com cada um dos operadores podem ser visualizadas no esquema apresentado na Tabela 1. Esta tabela apresenta as máscaras que devem ser aplicadas para a identificação de bordas verticais, bem como para a identificação de bordas horizontais em uma imagem, de acordo com os operadores de Roberts, Prewitt e Sobel.



Tabela 1. Máscara de operadores para detecção de bordas.

Fonte: Adaptado de Marques Filho e Vieira Neto (1999).

Operador	Máscara para detectar bordas verticais	Máscara para detectar bordas horizontais
Roberts	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$
Prewitt	$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$	$\begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$
Sobel	$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$	$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$

A detecção de bordas consiste basicamente na aplicação do operador (ou máscara) mediante um processo denominado de convolução que percorre cada um dos *pixels* da imagem na busca destas descontinuidades. A Figura 14 exhibe a aplicação das máscaras de Sobel (14b) e Prewitt (14c) sobre uma imagem em tons de cinza (14a), com o propósito de detectar bordas horizontais e verticais na imagem.

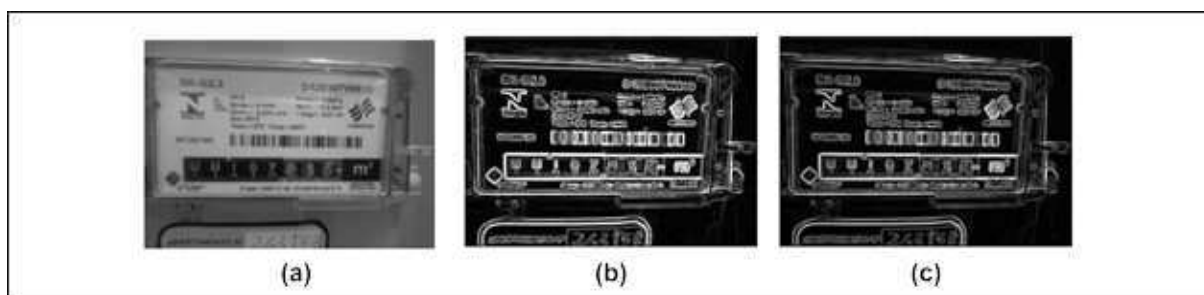


Figura 14. Detecção de bordas horizontais e verticais na imagem. (a) Imagem original; (b) Operador de Sobel; (c) Operador de Prewitt.

### 2.6.3 Projeção Horizontal e Vertical

De acordo com Pedrini e Schwartz (2008), as projeções horizontais e verticais de uma imagem binária consistem, respectivamente, na soma dos *pixels* em cada

linha da imagem (projeção horizontal), bem como na soma dos *pixels* em cada coluna (projeção vertical).

Analisando-se esta definição, nota-se que a projeção horizontal e vertical dos *pixels* de uma imagem pode ter o seu cálculo obtido com base em seus respectivos histogramas. A projeção vertical da imagem pode ser utilizada na tentativa de delimitar a largura de uma região de interesse, enquanto que a projeção horizontal pode ser utilizada para obtenção da altura da região.

#### 2.6.4 Morfologia Matemática Binária

Segundo Gonzalez e Woods (2002), a linguagem da morfologia matemática é a teoria dos conjuntos, e oferece uma abordagem unificada e poderosa para inúmeros problemas de processamento de imagens. O emprego da morfologia matemática possibilita a extração de componentes úteis para representar e descrever formas de regiões na imagem como, fronteiras, esqueletos e o fecho convexo. Pode também ser aplicada em técnicas de pré-processamento e pós-processamento, como filtragem morfológica, afinamento e poda.

A morfologia matemática consiste basicamente em extrair as informações relacionadas à geometria e à topologia de elementos que compõem uma imagem. Isto é feito transformando-se a imagem por meio de um conjunto bem definido denominado elemento estruturante (EE), que são menores que a imagem e podem possuir diversas formas como, por exemplo, a forma quadrada, circular ou em linha.

São considerados operadores básicos da morfologia matemática a erosão, a dilatação, a abertura e o fechamento que podem ser aplicados em imagens binárias bem como em imagens em níveis de cinza (SOILLE, 2003).

Nas próximas seções são descritos os operadores elementares da morfologia matemática.

#### 2.6.4.1 Operador de Erosão

No caso da aplicação da erosão em imagens binárias, o seu efeito será o de erodir as fronteiras do objeto, fazendo com que a área do objeto sofra uma redução de tamanho, enquanto que os buracos localizados dentro do objeto se tornem maiores.

A erosão é definida pela equação (13), na qual  $A$  é a imagem de entrada e  $B$  o elemento estruturante.

$$A \ominus B = \{x : Bx \subset A\} \quad (13)$$

#### 2.6.4.2 Operador de Dilatação

Na dilatação de imagens binárias, o efeito é de aumento gradual das fronteiras do objeto, proporcionando o crescimento em tamanho da área do objeto, enquanto que os buracos internos do objeto tornam-se menores.

A dilatação é definida pela equação (14), onde  $A$  é a imagem de entrada e  $B$  o elemento estruturante.

$$A \oplus B = \{x \in A : Bx \cap x \neq \emptyset\} \quad (14)$$

#### 2.6.4.3 Operador de Abertura

A abertura é geralmente utilizada para suavizar contornos em objetos, bem como eliminar contornos estreitos entre objetos, buscando também remover saliências finas. Para obtenção do efeito da operação de abertura aplica-se um operador de erosão seguido de um operador de dilatação no objeto. A abertura é definida pela equação (15), onde  $A$  é a imagem de entrada e  $B$  o elemento estruturante.

$$A \circ B = (A \ominus B) \oplus B \quad (15)$$

#### 2.6.4.4 Operador de Fechamento

A operação de fechamento é aplicada para remover pequenos buracos em objetos, preencher lacunas em contornos, bem como fundir separações estreitas entre objetos. Seu efeito é conseguido pela aplicação de um operador de dilatação seguido de um operador de erosão.

O fechamento é definido pela equação (16), onde  $A$  é a imagem de entrada e  $B$  o elemento estruturante.

$$A \bullet B = (A \oplus B) \ominus B \quad (16)$$

#### 2.6.4.5 Operadores *Top-hat* e *Bottom-hat*

As operações *top-hat* e *bottom-hat* são geralmente utilizadas para remoção de objetos da imagem. A primeira destaca regiões claras sobre fundos escuros e a segunda destaca regiões escuras sobre fundos claros. O efeito da operação de *top-hat* é obtido por meio da subtração da imagem original pelo resultado da operação de sua abertura. Para obtenção do efeito da operação *bottom-hat* subtrai-se a imagem original do resultado de seu fechamento.

#### 2.6.4.6 Operador de Gradiente Morfológico

O operador de gradiente morfológico destaca as transições entre os níveis de cinza destacando os contornos da imagem. O resultado da operação depende do EE utilizado, por exemplo, um EE pequeno é o elemento ideal para que os contornos sejam apresentados de forma bem definida e delgada. O resultado do gradiente morfológico é obtido pela diferença entre a imagem dilatada e a imagem erodida pelo mesmo EE (DOUGHERTY, 1992).

### 2.6.5 Análise de Componentes Conexos

Segundo Pedrini e Schwartz (2008), para verificar se dois *pixels* são conexos é necessário verificar se eles são vizinhos e se satisfazem determinados critérios de similaridade, tais como o nível de cinza.

A análise de componentes conexos é feita por meio da rotulagem de componentes. Busca-se transformar uma imagem binária até a obtenção de uma imagem simbólica que represente os objetos conectados na imagem original.

Basicamente o procedimento para rotulagem é o de examinar cada *pixel* juntamente com os seus vizinhos e aplicar um *label* (rótulo) a todos os *pixels* que correspondam ao mesmo componente.

A Figura 15 ilustra o processo de rotulagem de componentes conexos, onde na Figura 15(a) é exibido os *pixels* da imagem binária, na Figura 15(b) a imagem simbólica com cada um dos *labels* identificados conforme a aplicação da análise de componentes conexos, e na Figura 15(c) a imagem com os componentes conexos em destaque.

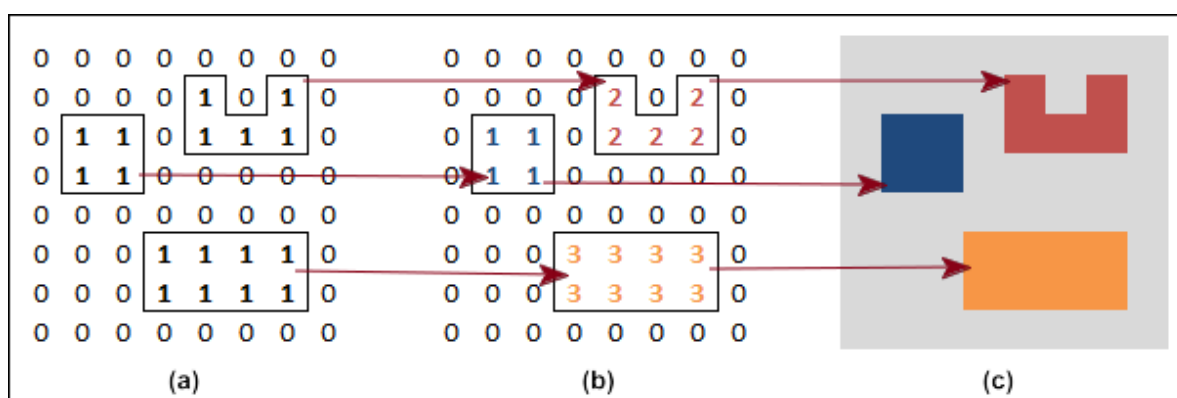


Figura 15. Rotulagem de componentes conexos.

O Quadro 1 apresenta os passos de uma abordagem possível para a rotulagem de componentes conexos.

#### Quadro 1. Passos para rotulagem de componentes conexos.

1	Varrer a imagem da esquerda para a direita e de cima para baixo.
---	--

2	<p>Se o <i>pixel</i> é igual a 1, então analise o vizinho de cima e o vizinho da esquerda.</p> <p>a) Se apenas um dos vizinhos possui <i>label</i>, copie o <i>label</i> para o <i>pixel</i> atual.</p> <p>b) Se ambos vizinhos possuem o mesmo <i>label</i>, copie o <i>label</i> para o <i>pixel</i> atual.</p> <p>c) Se os <i>labels</i> de ambos forem diferentes, copie o <i>label</i> do vizinho superior juntamente com um novo <i>label</i> para uma tabela de equivalência.</p> <p>d) Caso contrário, atribua um novo <i>label</i> ao <i>pixel</i> em análise.</p>
3	Se ainda restam <i>pixels</i> , retorne ao passo 2.
4	Varrer novamente a imagem, substituindo cada <i>label</i> pelo <i>label</i> de valor inferior encontrado na tabela de equivalência.

### 2.6.6 Interpolação Bilinear

Métodos de interpolação são empregados em imagens para transformá-las com o objetivo de proporcionar uma ampliação (*zooming*), encolhimento (*shrinking*), rotação ou correções geométricas utilizando de informações conhecidas para estimar valores em locais desconhecidos.

Segundo Gonzalez e Woods (2002), um método que contempla boa estimativa e baixo custo computacional é a interpolação bilinear. O método consiste em realizar a interpolação de valores dos 4 vizinhos mais próximos, calculando a média ponderada de distância destes vizinhos para determinar a intensidade de cada *pixel* ( $x', y'$ ) na imagem transformada, conforme ilustra a Figura 16.

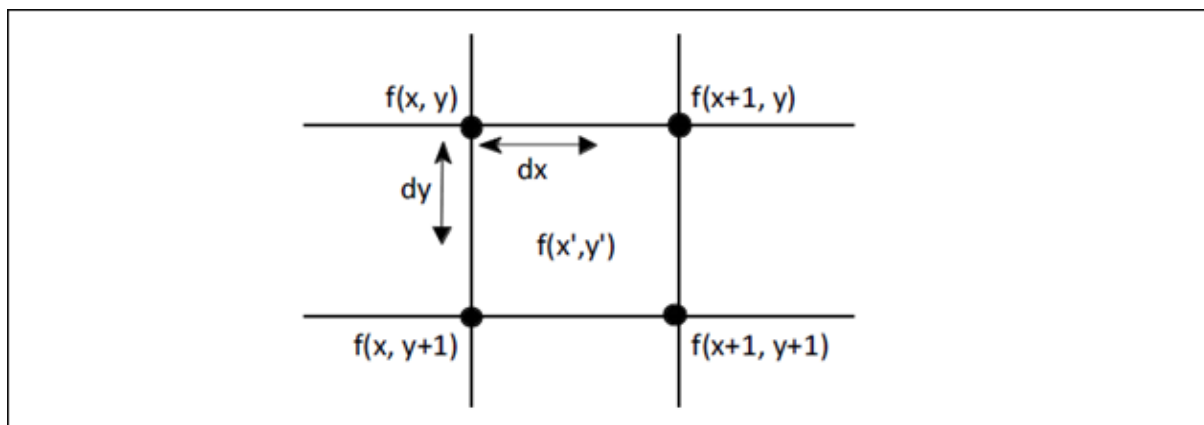


Figura 16. Interpolação bilinear.

Fonte: Pedrini e Schwartz (2008).

## 2.7 REPRESENTAÇÃO E DESCRIÇÃO

Um aspecto básico apresentado no desenvolvimento de sistemas que analisam imagens é a seleção de um conjunto de características relacionadas ao objeto de interesse, e que possam ser utilizadas para efeito de classificação destas imagens (PEDRINI e SCHWARTZ, 2008). Este conjunto de informações irá de alguma forma representar ou descrever determinada imagem.

Como exemplos de características é possível citar a ocorrência de dígitos de traços horizontais e verticais nas imagens, ou a presença de furos ou concavidades. A imagem do dígito 1 apresenta a característica de possuir um traço vertical, já nos dígitos 8 e 3 ocorrem furos e concavidades respectivamente.

Diversos métodos para extração de característica têm sido propostos na literatura, algumas destas abordagens são destacadas nas próximas seções.

### 2.7.1 Características Estruturais via Projeções

A extração de características estruturais de um objeto pode ser feita por meio do cálculo do perfil de projeções de *pixels* apresentados na imagem como a projeção horizontal e a vertical. A projeção horizontal de *pixels* é calculada pela soma dos *pixels* referentes ao objeto distribuídos em cada linha da imagem. A projeção vertical é obtida pela soma dos *pixels* do objeto distribuídos por coluna.

Depois de calculados os perfis, estes podem ser combinados para formar um único vetor de características.

### 2.7.2 Zoning

A extração de características utilizando a técnica de *Zoning* (PITAS, 2000), consiste em dividir a imagem em zonas retangulares de igual tamanho (ou próximo disto, no caso de imagens com altura ou largura que não permitam a divisão em áreas exatas).

Com as zonas delimitadas, toma-se uma medida que seja referente aos *pixels* pertencentes a cada uma destas zonas. Um exemplo de medida poderia ser o valor médio da intensidade dos *pixels* que compõem determinada zona, ou para o caso de imagens binárias o percentual de *pixels* distribuídos por zona que representem o objeto de interesse. Calculadas as medidas por zona, estas podem ser combinadas formando um único vetor de características.

### 2.7.3 Vetor de Comportamento

A técnica de extração de características baseada em vetor de comportamento (RUDEK, 1999), também leva em consideração a separação da área da imagem em zonas (ou subáreas). Para cada subárea da imagem determina-se a cor (ou o tom de cinza) que possua a maior frequência dentro da subárea.

Tendo extraído os valores de intensidades de *pixel* mais significativos em cada zona, organizam-se estes valores em um vetor, denominado pelo autor como vetor de comportamento. O vetor é obtido agrupando os valores encontrados em cada subárea na direção da esquerda para a direita e de cima para baixo. A Figura 17(a) exibe a divisão da imagem em zonas, em 17(b) exibe a identificação do nível de cinza mais frequente em cada zona.



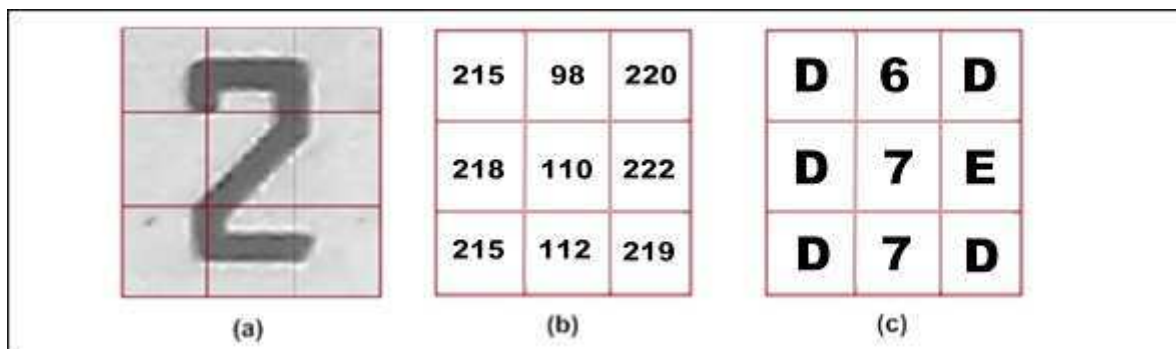


Figura 17. Esquema representativo do vetor de comportamento. (a) Imagem dividida em zonas; (b) Nível de cinza de maior frequência em cada zona; (c) Classificação de cada zona com base na tabela de intervalos de distribuição.

Fonte: Adaptado de Rudek (1999).

Determinados os valores mais frequentes, estes são classificados de acordo com o intervalo de distribuição apresentado na Tabela 2. As classes resultantes de cada classificação compõem assim o vetor de características com base na técnica de vetor de comportamento, cujos valores são apresentados na figura 17(c).

Tabela 2. Classes de distribuição do vetor de comportamento.

Classe	Intervalo de distribuição (maior intensidade dos <i>pixels</i> por zona)
1	000 a 016
2	017 a 033
3	034 a 050
4	051 a 067
5	068 a 084
6	085 a 101
7	102 a 118
8	119 a 135
9	136 a 152
A	153 a 169
B	170 a 186
C	187 a 203
D	204 a 220
E	221 a 237
F	238 a 255

## 2.8 TÉCNICAS PARA RECONHECIMENTO DE PADRÕES E CLASSIFICAÇÃO

A etapa de reconhecimento de padrões busca, partindo de um conjunto de medições, identificar objetos presentes na imagem que foram isolados pela etapa de segmentação. Cada objeto apresenta um padrão e as medições são baseadas nas características desse padrão. São considerados pertencentes à mesma classe de padrões, objetos que possuam uma ou mais características semelhantes. Nas próximas seções são descritas algumas técnicas que podem ser utilizadas para o reconhecimento de padrões.

### 2.8.1 K-vizinhos mais Próximos

A classificação utilizando o método dos K-vizinhos mais próximos (*KNN – K-Nearest Neighbours*) é uma técnica para classificar padrões que consiste em atribuir uma classe a um elemento desconhecido usando a classe da maioria de seus  $k$  vizinhos mais próximos, segundo uma determinada distância (no espaço de atributos).

A ideia deste classificador é basicamente a seguinte. A base de treinamento rotulada é a base de conhecimento do classificador. Neste caso, podemos dizer que o classificador é treinado com os dados rotulados que servirão de referência na tomada de decisão. Uma vez que o classificador está projetado, ou seja, possui sua base de treinamento rotulada, o critério de decisão utilizado para obter qual é a classe de uma amostra desconhecida é o seguinte:

- 1) Dada a amostra desconhecida, calcule a sua distância a todas as amostras da base de treinamento e selecione as  $k$  amostras mais próximas.

- 2) Dentre as  $k$  amostras mais próximas, escolha a classe que seja maioria entre elas. Se houver empate, utilize a classe mais próxima das maiorias empatadas.

O classificador *k-nn* possui três passos principais:

- 1) Cálculo da distância entre duas amostras quaisquer.

2) A ordenação das distâncias para identificar as  $k$  amostras mais próximas.

3) A identificação da maioria dos rótulos das  $k$  amostras mais próximas.

Para buscar os  $k$ -vizinhos mais próximos entre a amostra desconhecida e as amostras da base de conhecimento, precisamos de uma função distância entre duas amostras quaisquer utilizando os atributos destas amostras. Estes atributos podem ser, por exemplo, valor médio da imagem, variância, percentis, histograma, estatísticas do gradiente entre outros.

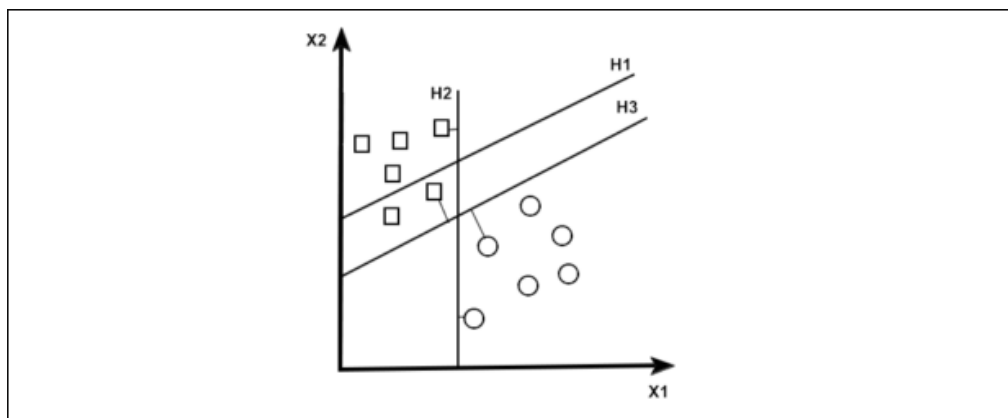
Para o cálculo da distância pode ser utilizada a distância euclidiana que é dada pela raiz quadrada da soma das diferenças quadráticas dos atributos.

### 2.8.2 *Support Vector Machine*

*Support Vector Machine* (SVM) é um método de aprendizagem de máquina desenvolvido por Vapnik (2013), cuja estratégia de aprendizagem é baseada em um algoritmo de Otimização Matemática e que implementa um limite derivado da Teoria de Aprendizagem Estatística.

O método propõe a criação de um hiperplano como superfície de decisão, que apresente a separação ideal entre duas classes, maximizando a distância entre os pontos mais próximos da classe e o hiperplano. Estes pontos são denominados elementos de borda, e os vetores de suporte são as distâncias entre os elementos de borda e o hiperplano.

A Figura 18 exibe um exemplo de classificação pelo método SVM compreendida por 3 hiperplanos, na qual o primeiro hiperplano (H1) não fornece a capacidade de separar de forma correta as duas classes. Já o segundo hiperplano (H2) consegue separar as classes, mas faz com menor margem (distância) entre os pontos e H2. Por fim, o terceiro hiperplano (H3) separa corretamente as classes e maximiza a margem entre os pontos e H3.



**Figura 18. Exemplo de classificação via método SVM.**

O treinamento do método se baseia na Teoria de Aprendizagem Estatística para supervisionar a capacidade e prevenir o efeito de super especialização, por meio do controle das medidas de margem do hiperplano. Tem também como base a Teoria de Otimização para prover técnicas matemáticas necessárias para encontrar o hiperplano otimizando essas medidas. Detalhes destas teorias podem ser encontrados no trabalho de Dos Santos (2002).

Para problemas cujas classes de dados não se apresentem linearmente separáveis, emprega-se o uso de funções de *kernel* (VAPNIK, 2013). O emprego destas funções será o responsável por mudar a dimensionalidade das classes de dados, as transformando em classes linearmente separáveis. As funções de *kernel* mais comuns são: hiperbólica, gaussiana, polinomial e linear.

Como vantagens o SVM apresenta a capacidade de trabalhar com um grande volume de amostras e a rapidez na classificação. Por outro lado, o tempo de treinamento pode ser demasiadamente longo dependendo do número de exemplos e dimensionalidade dos dados.

### 2.8.3 Redes Neurais Artificiais

Uma rede neural artificial (RNA) é um modelo de grafo orientado em que os nós representam neurônios artificiais e as arestas orientadas denotam as conexões existentes entre as entradas e saídas do neurônio (PEDRINI e SCHWARTZ, 2008). O modelo mais simplificado de RNA é composto de um único neurônio denominado *perceptron*. Este modelo mapeia múltiplas entradas com valores reais para uma

saída que é obtida mediante uma função de ativação, indicando por meio de um valor binário a ativação ou não de determinado estado.

A base do processamento de uma RNA é o neurônio, o qual possui três elementos básicos: conjuntos de sinapses, ou elos de conexão, onde são calculados e armazenados os pesos sinápticos da RNA, o somador que é responsável pelo somatório dos pesos multiplicados pelos valores de entrada da RNA e a função de ativação que limita a amplitude de sinal de saída de um neurônio. A Figura 19 ilustra a estrutura de um neurônio artificial.

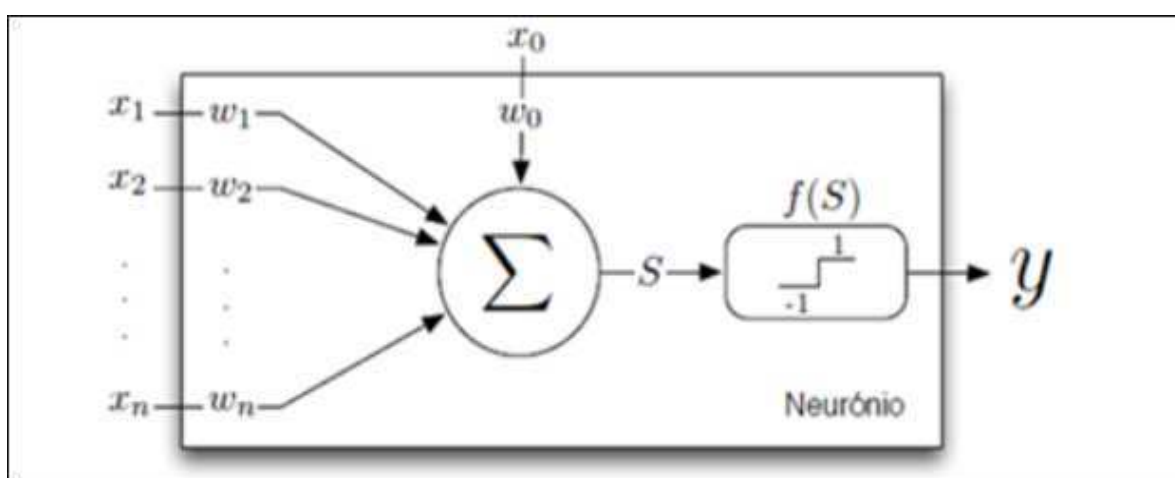


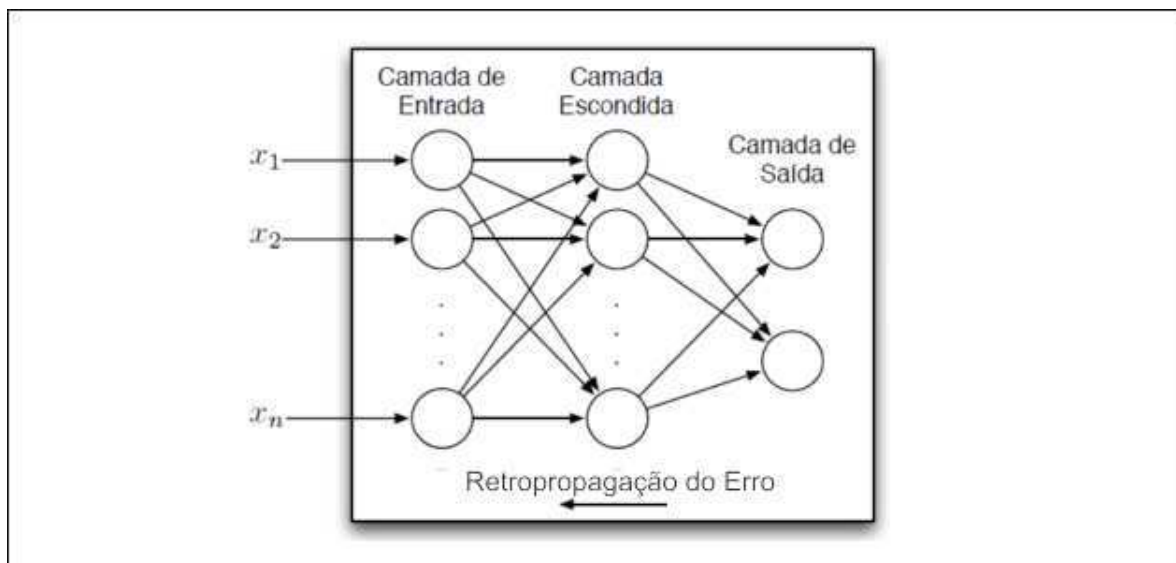
Figura 19. Estrutura de um neurônio artificial.

Fonte: Adaptado de Pedrini e Schwartz (2008).

Na Figura 19 observam-se os valores de entrada da rede representados por  $x_1, x_2, \dots, x_n$ , os pesos sinápticos representados por  $w_1, w_2, \dots, w_n$ , o somador juntamente com o *bias* (viés)  $x_0$  e  $w_0$ , e a saída linear  $S$  que passa por uma função de ativação  $f(S)$  e que resulta no valor de saída  $y$  do neurônio.

O modelo de um único neurônio é capaz apenas de classificar padrões linearmente separáveis, para a resolução de problemas não lineares são utilizadas RNAs de múltiplas camadas.

A Figura 20 ilustra uma RNA de múltiplas camadas, na qual podemos observar a presença de camadas de entrada, oculta e de saída.



**Figura 20. Estrutura de uma RNA de múltiplas camadas.**

**Fonte: Adaptado de Pedrini e Schwartz (2008).**

O processo de aprendizado de uma rede neural artificial acontece, basicamente, de duas formas: aprendizado supervisionado e aprendizado não supervisionado.

No aprendizado não supervisionado, nenhum conjunto de pares de entrada e saída é apresentado à rede. Neste caso são fornecidas condições para realizar uma medida, independente da tarefa ou da qualidade da representação, cuja rede deve aprender. Já o aprendizado supervisionado insere em sua estrutura uma espécie de instrutor que confere o quanto a rede está próxima de uma solução aceitável, adaptando na concepção do treinamento os pesos entre os neurônios, de modo a prover uma menor diferença entre as saídas desejadas e a obtida.

Para o caso do aprendizado supervisionado, um algoritmo de treinamento bastante conhecido na literatura é o algoritmo *Back-propagation* (BP). Ele é constituído por duas etapas: a propagação e a retropropagação de um conjunto de sinais através da rede.

A propagação consiste na aquisição dos dados pela camada de entrada e sua propagação por toda rede, produzindo uma saída. A saída da rede neural é comparada com a saída desejada e um valor de erro é calculado. A partir daí começa o processo de retropropagação, no qual esse erro é propagado de volta à

rede neural e usado para ajustar os pesos. Busca-se reduzir o erro a cada iteração para que o resultado aproxime-se cada vez mais da saída desejada.

#### 2.8.4 *Extreme Learning Machine*

O método de aprendizagem de máquina *Extreme Learning Machine* (ELM), proposto por (HUANG; ZHU; SIEW, 2004) é derivado da visão atual de redes neurais como aproximadores universais. Utiliza redes *Single-hidden Layer Feedforward Network* - SLFN, que por apresentarem uma única camada oculta, são consideradas um caso especial das redes *Multilayer Perceptron* - MLP (SILVESTRE, 2015).

A ELM se destaca devido ao baixo custo computacional para treinamento da rede. Sua essência está no uso de pesos sinápticos aleatórios na camada oculta. Deste tipo de configuração origina-se um modelo linear para os pesos sinápticos da camada de saída da rede. Estes pesos são calculados de forma analítica por meio de uma solução de mínimos quadrados (HUANG; WANG; LAN, 2011).

Uma das principais dificuldades em se treinar redes neurais artificiais do tipo MLP é o ajuste pesos e parâmetros (taxa de aprendizagem, número de iterações, critérios de parada). Em geral, abordagens tradicionais realizam a retropropagação do erro de treinamento através das camadas da rede. O algoritmo *Backpropagation* (BP) é utilizado para realizar esse tipo de treinamento, que por ser baseado no método do gradiente descendente geralmente apresenta lentidão no seu processo de aprendizagem, além da possibilidade de ficar presos em mínimos locais.

O método ELM visa dispensar a necessidade do ajuste manual destes parâmetros, bem como busca resolver o problema dos mínimos locais em relação ao algoritmo BP, haja visto que o método ELM não é baseado nas teorias de gradiente descendente. O método prima por tornar mais rápido o treinamento da rede, sem perder a capacidade de resolver corretamente os mesmos problemas. A Figura 21 apresenta a estrutura de uma rede neural ELM, onde:

- 1)  $\mathbf{x}$  é constituído por uma matriz com  $\mathbf{N}$  vetores de treinamento compostos por  $\mathbf{d}$  características, e representa os dados de entrada da rede.

- 2)  $\mathbf{a}$  representa os pesos aleatórios dos  $L$  neurônios da camada escondida, que são fatores de multiplicação dos dados de entrada.
- 3)  $\mathbf{b}$  representa o *bias*, um valor constante que é o responsável pelo aumento ou diminuição da entrada da função de ativação.
- 4) A quantidade de neurônios apresentada na camada de entrada  $d$  deve corresponder a quantidade de características apresentada no vetor de treinamento.
- 5) A quantidade de neurônios da camada escondida  $L$  é determinada por meio de experimentos.

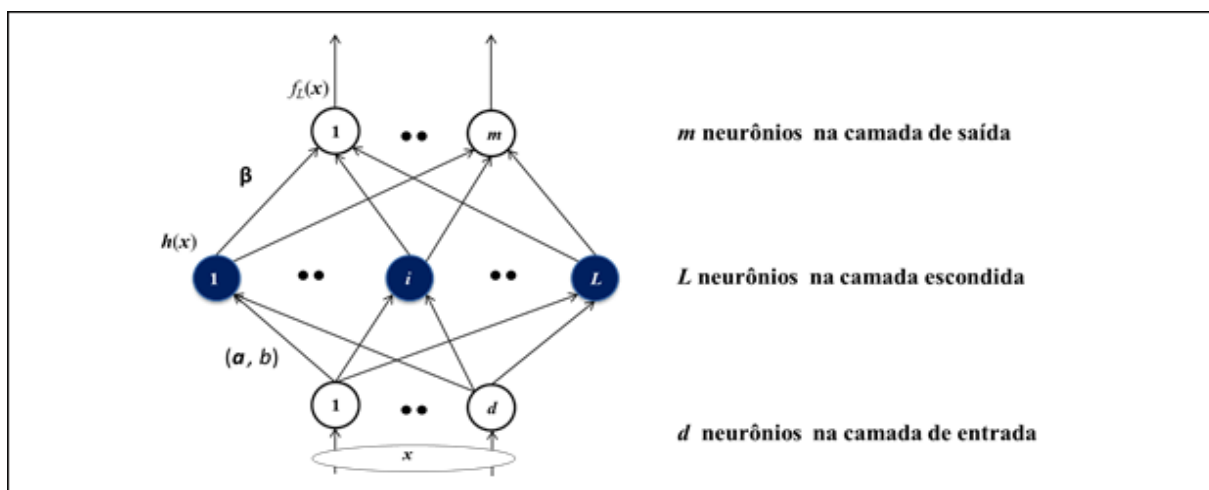


Figura 21. Arquitetura da rede neural ELM.

Fonte: Boaretto (2007).

A regra geral para o processo de aprendizagem dos neurônios de uma ELM é definida pela equação (17).

$$\sum_{i=1}^H \beta_i g(W_i X_j + b_i) = t_j; \quad j = 1, \dots, N \quad (17)$$

Aonde:

- $H$  é a quantidade de neurônios na camada oculta;
- $N$  é o número de vetores de treinamento;
- O vetor  $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$  possui os pesos da camada oculta;
- O vetor  $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$  possui os pesos da camada de saída;



No processo de treinamento para cada entrada  $\mathbf{x}_j$  de dimensão  $n$ , apresenta-se o valor conhecido  $\mathbf{t}_j$  de dimensão  $m$ . Uma função de ativação diferenciável  $g(x)$ , em geral do tipo logística, é escolhida para regular assintoticamente a saída da rede ELM.

De acordo com Huang, Zhu e Siew (2006) ao se treinar uma *SLFN* com  $K$  neurônios na camada oculta e função de ativação  $g(x)$ , com as entradas aprendendo através de  $N$  amostras distintas  $(x_i, t_i)$ , sendo  $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbb{R}^n$  e  $\mathbf{t}_i = [t_{i1}, t_{i2}, \dots, t_{in}]^T \in \mathbb{R}^m$ , gerando-se aleatoriamente os pesos de entradas e o bias da camada oculta, o sistema não-linear é convertido para um sistema linear descrito por  $\mathbf{H}\boldsymbol{\beta} = \mathbf{T}$ , onde:

$\mathbf{H} = h_{ij}(i = 1, \dots, N \text{ e } j = 1, \dots, K)$  é a matriz de saída da camada oculta;

$h_{ij} = g(w_j \cdot x_i + b_j)$  é a saída do  $j$ -ésimo neurônio oculto relacionado a  $x_i$ ;

$\mathbf{w}_j = [w_{j1}, w_{j2}, \dots, w_{jn}]^T$  é o vetor de pesos das conexões do  $j$ -ésimo neurônio da camada oculta e os neurônios de entrada;

$b_j$  se refere ao bias de  $j$ -ésimo neurônio da camada oculta;

$\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{iK}]^T$  denota o vetor de pesos que conectam o  $j$ -ésimo neurônio da camada oculta e os neurônios da camada de saída da rede;

$\mathbf{T} = [t_1, t_2, \dots, t_n]^T$  é o vetor referente às saídas desejadas.

Os valores dos pesos entre a camada oculta e a camada de saída podem então ser obtidos por meio do método de mínimos quadrados, cuja norma mínima é dada por  $\boldsymbol{\beta}_i = \mathbf{H}^\dagger \mathbf{T}$ , aonde  $\mathbf{H}^\dagger$  é a pseudoinversa de Moore-Penrose da matriz  $\mathbf{H}$  (SERRE, 2002).

É possível provar que a rede ELM independe dos pesos da camada oculta  $\mathbf{w}_j$  os quais podem ser inicializados com valores aleatórios (HUANG, ZHU; SIEW, 2006). Após o treinamento da rede ELM, são apresentados dados desconhecidos para obter uma predição da classe à qual esses dados mais se assemelham.

### 2.8.5 Avaliação de Métodos de Classificação

A avaliação do desempenho de algoritmos de classificação deve levar em conta a sua capacidade de generalização com base em amostras que não foram utilizadas para construção do classificador. O conjunto total de amostras disponíveis deve ser dividido em dois grupos: um primeiro grupo de treino que será utilizado pelo algoritmo para induzir aprendizagem ao classificador, e outro grupo de teste que deve ser utilizado para verificar o desempenho do classificador em relação ao índice de erros e acertos. Nas próximas seções são descritos avaliadores que podem ser utilizados para este propósito.

#### 2.8.5.1 Método de Validação *Holdout*

Este método consiste em dividir o conjunto de amostras em dois subconjuntos: um conjunto de treino contendo  $2/3$  (67%) das amostras e outro conjunto de testes contendo o  $1/3$  (33%) das amostras restantes. Tem como principal vantagem o tempo de processamento, porém dependendo do tamanho da base de amostras e de como é distribuída a composição dos subconjuntos a avaliação pode apresentar alta variância em seu resultado.

#### 2.8.5.2 Método de Validação Cruzada *K-fold*

Este método é uma generalização do método *Holdout* e consiste em aplicar no conjunto de amostras sucessivas partições, combinando-as entre conjunto de treino e teste (STONE, 1974).

As amostras são divididas em  $k$  partes, todas de igual tamanho, sendo comum  $k=10$  o que determina que nove partes sejam estabelecidas para o conjunto de treinamento e a parte restante selecionada para validar o classificador. Este conjunto de validação é alternado em cada passo do processo que neste caso será executado 10 vezes. A taxa de avaliação é calculada com base na média das taxas de erro obtidas em cada iteração.

### 2.8.5.3 Método de Validação *Leave-one-out*

Este é um caso particular do método de validação cruzada, que atribui ao valor de  $k$  a quantidade total de amostras do conjunto inicial (LUNTS; BRAILOVSKY, 1967). Os subconjuntos de teste são formados por apenas uma amostra e o conjunto de treino formado pelas amostras restantes.

A validação cruzada *Leave-one-out*, é um método que possui um custo computacional alto. Sua utilização é recomendada apenas em situações na qual o conjunto de amostras é pequeno.

## 2.9 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados os fundamentos com conceitos básicos importantes relacionados aos processos de segmentação, extração de características e reconhecimento de padrões em imagens. Estes fundamentos serão oportunos para a compreensão do método proposto neste trabalho relacionado ao problema do reconhecimento de dígitos apresentados em medidores de consumo de gás natural. O estado da arte relacionado a este problema é apresentado no próximo capítulo.

### 3 ESTADO DA ARTE

#### 3.1 INTRODUÇÃO

Apesar da complexidade apresentada, sistemas que utilizam técnicas de processamento digital de imagens e reconhecimento de padrões para a tarefa de reconhecimento de textos, podem encontrar possibilidades variadas de aplicações. Alguns exemplos dessas aplicações encontradas na literatura são: no reconhecimento de placas de automóveis apresentado no trabalho de Gazcón, Chesñevar e Castro (2012) e Wen et al. (2011), ou na captura e reconhecimento de texto em cenas naturais do cotidiano descrito em Epshtein, Ofek e Wexler (2010) e Zhang et al. (2002). Também são encontradas aplicações para reconhecimento de textos em objetos específicos como no caso de rótulos de garrafas de bebidas descrito em Arrighi et al. (2012) ou etiquetas apresentadas em embalagens descrito em Lim et al. (2009).

A literatura apresenta também trabalhos que abordam o emprego de técnicas de processamento digital de imagens e reconhecimento de padrões para reconhecer dígitos apresentados em medidores de energia. Abordagens distintas são mencionadas no que diz respeito aos algoritmos empregados em cada trabalho.

#### 3.2 RECONHECIMENTO DE DÍGITOS APRESENTADOS EM MEDIDORES

Alexandria et al. (2004) apresenta uma abordagem para reconhecimento de algarismos em medidores de energia elétrica. Um protótipo é criado em ambiente PC com câmera acoplada para captura da imagem do medidor. A segmentação utilizando a limiarização é empregada para separar os algarismos na imagem capturada. A imagem de cada algarismo é normalizada para uma resolução de 16 x 16 *pixels* e o seu reconhecimento fica a cargo de uma RNA com topologia de 256 entradas, 64 neurônios em camada intermediária e 10 sinais de saída que representam os dígitos de 0 a 9. Para o treinamento da RNA é utilizado o algoritmo BP.

Castells-Rufas e Carrabina (2006) utilizam um sensor CMOS em conjunto com um dispositivo FPGA acoplados ao medidor para captura da imagem. Após a captura a segmentação é feita pela aplicação de uma limiarização (*threshold*) simples que consiste em encontrar um limiar composto pela média entre o menor e o maior valor de *pixel* na imagem. Feita a segmentação para cada dígito encontrado, a extração de atributos é obtida pela geração de uma sequência de símbolos que representa um padrão observado na análise de cada linha de *pixels* da imagem. A geração destes símbolos é feita por uma máquina de estados (FSM, do inglês *Finite State Machines*) alimentada com cada linha de *pixels* resultando em um vetor de características para cada dígito segmentado. A classificação dos dígitos é realizada pela comparação com um modelo de caracteres numéricos obtidos por um processo de treinamento. Por fim, a comparação da sequência de símbolos com o modelo é feita por programação dinâmica utilizando o algoritmo de Needleman e Wunsch (1970) por meio do alinhamento de sequências. O desempenho da aplicação desta abordagem em uma base de dados contendo 396 imagens alcançou 99,46% de taxa de sucesso.

O trabalho de Shu, Ma e Jing (2007) aborda o reconhecimento dos dígitos em medidores de energia elétrica, cuja imagem é adquirida em tons de cinza por uma câmera de vídeo acoplada ao medidor. Nesta imagem é aplicada uma limiarização para transformar a imagem em preto e branco. Para identificar e segmentar a área de interesse no medidor, os autores partem da premissa que, como a câmera é fixa, a área contendo os dígitos apresenta as características de ser retangular e de tamanho padrão, com os dígitos após o processo de limiarização se apresentando na cor branca sobre fundo preto.

Baseados nestas características os autores aplicam a técnica de projeção horizontal dos *pixels* da imagem buscando encontrar o intervalo de linhas no qual a área contendo os dígitos deverá estar localizada. A análise desta projeção indica que a área deve estar localizada entre as linhas 400 e 600 da imagem. Na sequência, partem do mesmo raciocínio para encontrar o intervalo de colunas no qual a área contendo os dígitos deverá estar posicionada, para isto utilizam a técnica de projeção vertical dos *pixels* da imagem. A análise desta projeção indica que a área contendo os dígitos encontra-se entre as colunas 250 e 380 da imagem.

Operadores morfológicos de erosão e dilatação são aplicados à imagem segmentada para eliminação de ruídos que possam prejudicar a segmentação individual de cada dígito. Para obter cada dígito de forma individual primeiramente as cores da imagem são invertidas de forma que os dígitos passam a ser pretos sobre fundo branco, em seguida aplica-se novamente a técnica de projeção vertical.

Para o processo de reconhecimento dos dígitos é feita uma análise baseada nas características dos traços de cada dígito, que segundo a análise podem possuir um aspecto reto em direção horizontal ou vertical, bem como um aspecto curvo de forma completa ou incompleta, além do posicionamento do traço que pode estar localizado no topo da imagem do dígito ou na base. Tomando como exemplo a característica curva, o dígito 8 possui duas curvas completas, já os dígitos 0, 4, 6 e 9 possuem apenas uma curva completa. Esta curva no dígito 6 está posicionada na base do dígito, enquanto que no dígito 9 está posicionada no topo. Já para os números 1, 2, 3, 5 e 7 a análise pode ser feita pelo traço reto, tendo, por exemplo, o dígito 1 classificado como apresentando um traço vertical, o dígito 2, um traço horizontal na base, e assim por diante. Desta forma os autores vão descrevendo as características de cada dígito em termos da apresentação de seus traços na imagem. A aplicação das técnicas de projeção horizontal em conjunto com a rotulagem de componentes conexos em cada imagem individual de dígito, trabalhando sobre cada característica extraída, forma a base do algoritmo para reconhecimento dos dígitos. Os autores não reportam resultados numéricos do experimento, no entanto relatam que o algoritmo apresentou um reconhecimento rápido e com alta precisão.

Na abordagem de Oliveira, Cruz e Bensebaa (2009) que trata do reconhecimento de dígitos em medidores de energia elétrica, as imagens são capturadas na dimensão 3056x2292 *pixels* utilizando uma câmera CCD. O aprimoramento da imagem na etapa de pré-processamento é feito por meio da aplicação de um filtro homomórfico a fim de compensar os ruídos ocasionados pela não uniformidade de iluminação. Na etapa de segmentação, os autores por meio de observação presumem que a região de interesse a ser segmentada apresenta-se entre a segunda e terceira linha horizontal de elementos na imagem. O processo de segmentação é realizado utilizando a limiarização pelo método de Otsu (1979) e na sequência é aplicado um operador morfológico de abertura para eliminação de

ruídos. Em seguida faz-se a detecção de bordas pelo operador de gradiente morfológico. Isto feito aplica-se um operador de abertura utilizando um elemento estruturante em linha com o objetivo de deixar apenas linhas horizontais, permitindo, assim, encontrar a segunda e terceira linha de grande proporção. De acordo com a interpretação visual dos autores, isto delimita a região de interesse que contém a área de leitura do medidor.

Após encontrar o espaço em termos da altura que delimita a região de interesse, utilizam um operador morfológico de abertura com elemento estruturante grande para encontrar o limite esquerdo e direito da região a ser segmentada. Este procedimento remove tanto os ruídos da imagem quanto os próprios dígitos da região. Busca-se então a primeira ocorrência de *pixel* branco à esquerda e a direita do centro da imagem para em seguida eliminar os *pixels* que não pertençam a este intervalo, tendo como resultado a área de interesse segmentada.

Para segmentação individual dos dígitos utilizou-se um operador de fechamento a fim de remover ruídos e ajustar bordas irregulares dos dígitos. Juntamente com esta técnica foi utilizada também uma técnica de extração de componentes conectados. Primeiramente foi efetuada uma análise da linha horizontal localizada no centro da altura da imagem, da esquerda para a direita, posição na qual os autores novamente estimam ser o local no qual possam ser encontrados os dígitos a serem segmentados. Quando um *pixel* branco é encontrado, utiliza-se uma função que extrai componentes conexos. Para a etapa de reconhecimento foi utilizado o algoritmo *KNN* (*K-Nearest Neighbors*). Os autores não destacaram resultados em termos de taxa de acerto no reconhecimento dos dígitos e nem mesmo em desempenho da tarefa, apenas citando que ao basearem-se em outros experimentos, notou-se que o algoritmo obteve alta precisão.

No trabalho de Zhao et al. (2009), é abordado o reconhecimento dos dígitos em medidores de energia elétrica e de gás. Para execução dos experimentos um ambiente em laboratório foi montado com utilização de um sensor CMOS em conjunto com um microprocessador DSP para a tarefa de aquisição das imagens. A imagem capturada é tratada utilizando a equalização de histograma para aprimorar o contraste e o brilho da imagem original. Para localizar a região de interesse é utilizada a segmentação por análise de projeção: o método da projeção horizontal é

utilizado para localizar o contador de consumo e o método de projeção vertical é utilizado para separar cada um dos dígitos apresentados no contador. Cada imagem de dígito individual é então padronizada para o tamanho 21x15 *pixels* sobre a qual é aplicada a técnica de limiarização.

Os dígitos são classificados por meio de uma rede neural artificial (RNA) utilizando o algoritmo BP para treinamento. A rede neural é definida em três camadas: a camada de entrada é constituída por 315 nós, a camada intermediária constituída por 10 nós e a camada de saída também constituída por 10 nós em correspondência ao número de cada dígito a ser reconhecido (0 a 9). Foram utilizadas 500 imagens no experimento e o algoritmo produzido apresentou taxa de reconhecimento na ordem de 95% com velocidade inferior a 9s para o reconhecimento dos dígitos.

Bai et al. (2010), também fazem uso de câmera acoplada ao medidor para aquisição da imagem. Na fase de pré-processamento utilizam a equalização de histograma como forma de distribuir os níveis de cinza da imagem de maneira uniforme melhorando assim o seu contraste. Ainda nesta fase é aplicado o filtro da mediana para eliminação de ruídos da imagem sem prejudicar as bordas na imagem, já que a preservação das bordas é de real importância para o posterior reconhecimento dos dígitos presentes no medidor. Para a localização da ROI a ser processada foi utilizada uma limiarização (*threshold*) global pela qual a imagem é binarizada. Os autores também utilizam as técnicas de projeção horizontal e vertical tanto para a detecção da ROI que possui o contador, quanto para a segmentação individual de cada dígito presente no contador. Estas projeções permitem a contagem dos *pixels* existentes em cada linha e coluna proporcionando a identificação dos limites da posição inicial e final de cada um dos dígitos apresentados na imagem do contador.

A extração de atributos é feita utilizando a técnica de momentos invariantes, capaz de representar propriedades da imagem invariantes com relação a aspectos como a rotação e translação. Para tal são aplicadas sete fórmulas também denominadas de momentos invariantes de Hu (1961). Para o reconhecimento foi utilizada uma RNA cuja topologia foi definida em 3 camadas, com a camada de entrada composta por 7 nós que correspondem aos sete momentos de Hu, uma



camada intermediária com 14 nós e uma camada de saída com 10 nós representando os dígitos de 0 a 9.

Para o treinamento da rede foi utilizado o algoritmo BP, treinado com imagens de 50 instrumentos medidores com diferentes níveis de ruído e alcançando uma taxa de 94% de acerto. Porém, observou-se que as condições para aquisição de imagens com qualidade eram favoráveis e em ambientes em que a qualidade da imagem não possa ser garantida, a taxa de precisão pode cair. Foi observado também que como a proposta não envolvia um sistema de tempo real, a velocidade da RNA não foi otimizada, propondo estas duas observações para trabalhos futuros.

Lei et al. (2010), apresentam um método para reconhecimento de dígitos em medidores de energia elétrica por meio da utilização de um dispositivo *handheld* PC. Após a captura da imagem pelo dispositivo é feita a transformação em tons de cinza e aplicado um filtro de média de vizinhança 3x3 para a redução de ruídos. Para encontrar a área que apresenta os dígitos os autores executam uma limiarização pelo método de Otsu (1979) combinado com operações morfológicas de erosão e dilatação. Tendo encontrado a área de interesse, ajusta-se a distorção proveniente de inclinação dos dígitos da área utilizando o método de interpolação bilinear para girar a imagem.

Para a segmentação de cada dígito individualmente são utilizadas técnicas de crescimento de regiões em conjunto com a esqueletização da imagem. O reconhecimento é baseado no esqueleto do dígito por meio da aplicação de um algoritmo que busca analisar cada *pixel*, baseado em sua vizinhança-8, para encontrar grupos de *pixels* que representem furos ou concavidades. Por exemplo, na ocorrência de dois furos identifica-se o dígito 8, na ocorrência de um furo identifica-se os dígitos 0, 4, 6 ou 9, quando a ocorrência for de concavidade, identifica os dígitos 2, 3, 5 ou 7. O dígito 1 é identificado baseado na proporção entre altura e largura. Foram utilizadas 850 imagens para validação do método de reconhecimento dos dígitos, os experimentos mostraram um índice de exatidão de 99,02% para o reconhecimento e uma taxa de erro de 0,98%.

No trabalho de Nava-Ortiz, Gómez e Diaz-Perez (2011), é tratado o reconhecimento de dígitos em imagens de medidores de consumo de água. As imagens primeiramente têm seu espaço de cores reduzido para 256 tons de cinza e

suas dimensões ajustadas para a resolução de 120x30 *pixels*. Do conjunto de imagens capturadas são escolhidas algumas imagens para a construção da base de treinamento do experimento, estas imagens são recortadas manualmente proporcionando a obtenção de imagens com dígitos únicos de 0 a 9 e resultando em um total de 2.240 imagens, todas com o tamanho de 19x30 *pixels*. Para o procedimento de teste do experimento foram capturadas 217 imagens que apresentam a exibição da leitura completa do medidor.

Para abordar a tarefa de segmentação, inicialmente é aplicada a limiarização por Bernsen (1986) descrita na seção 2.6.1. O resultado da aplicação da limiarização não identifica totalmente a área de interesse devido à presença de ruídos que ainda dificultam a identificação dos números no medidor. A abordagem para a eliminação destes ruídos parte da premissa dos autores de que:

- 1) A área da imagem do medidor que contém a região de interesse é maior que qualquer outro objeto na imagem.
- 2) A região de interesse está centrada ao redor do centroide da imagem.

Desta forma, a estratégia dos autores para eliminar os ruídos apresentados após o processo de binarização envolve a rotulagem da imagem utilizando conectividade-4 e a obtenção do centroide da imagem. A região correspondente aos dígitos é então obtida executando-se um cálculo que envolve o valor da área de cada região rotulada, e a média da distância Euclidiana do centro da imagem para todos os *pixels* dentro de cada região rotulada. Leva-se em consideração também a máxima distância do objeto rotulado para o centroide da imagem.

Para superar a dificuldade no processo de extração de características das imagens, devido aos diferentes tamanhos apresentados para cada dígito, o tamanho de cada imagem com o dígito segmentado é ajustado para 20x20 *pixels*. A extração de características é feita utilizando a distribuição estatística dos *pixels*, baseada nas características de cruzamentos e distâncias, momentos centrais normalizados e zoneamento.

Para o reconhecimento dos dígitos são aplicados e comparados dois classificadores de distância mínima, baseado na distância Euclidiana e na distância de Mahalanobis. O primeiro foi treinado para determinar o centroide das classes

enquanto que o segundo treinado para definir o centroide e a matriz de covariância das classes. O treinamento foi executado em um microcomputador PC tendo como entrada a base de características extraídas empregando os métodos estatísticos de distribuição dos *pixels*. Os testes foram executados utilizando um smartphone, e avaliou tanto o reconhecimento da base de dígitos individuais quanto da base com as imagens de leituras completas. Combinaram-se os dois classificadores com cada base de características extraídas por cada um dos três métodos estatísticos individualmente, bem como com os três métodos em conjunto. O resultado apontou para a utilização combinada das três bases de características em conjunto com o classificador baseado na distância de Mahalanobis, com taxa de reconhecimento na ordem de 98,9% para a base de dígitos separados e 93,1% para a base com as imagens de leituras completas.

No trabalho de Nodari e Gallo (2011) é abordada a detecção e segmentação dos dígitos de um medidor sem a utilização de abordagens tradicionais de pré-processamento da imagem. Para tanto, os autores utilizam uma árvore de redes neurais denominada *MultiNOD*. Para a configuração da árvore considerou-se a seguinte estratégia, nós pertencentes ao primeiro nível da árvore foram usados para determinar a região de interesse na imagem e os subsequentes utilizados para eliminar os falsos positivos relacionados a ruídos. O *MultiNOD* foi aplicado em duas fases, na primeira para detectar a região correspondente ao contador no medidor. Foi utilizado um banco de 102 imagens de treinamento e 51 imagens de teste. Na segunda fase, para detectar os dígitos separadamente, foi construído um novo banco baseado no recorte de imagens do banco de imagens anterior consistindo em 47 imagens de treino e o mesmo número de imagens de teste. Embora a taxa de precisão de detecção do fundo da imagem tenha alcançado precisão acima de 90% em ambos os casos, a detecção da região girou em torno de uma taxa de precisão não maior que 70%, o que demonstra que a abordagem não se tornou tão robusta. Além do mais, para o reconhecimento dos dígitos utilizou-se um OCR comercial padrão denominado *Tesseract* obtendo uma taxa de reconhecimento que se julgou insuficiente (58,73%), o que veio a destacar a complexidade da tarefa de OCR neste domínio de problema, isto devido à particularidade relacionada a ruídos apresentados nas imagens capturadas.

Jabba et al. (2012), propõem um algoritmo para o reconhecimento de dígitos que é validado em um modelo específico de medidor de energia elétrica. O experimento foi realizado em laboratório com a aquisição das imagens feitas por uma câmera *wireless* localizada a 5,5 cm do medidor, em seguida envia-se para um micro PC localizado em laboratório as imagens capturadas. Esta câmera é calibrada para capturar as imagens com equilíbrio de brilho e contraste. A abordagem para segmentar a imagem é feita utilizando a análise do espaço de cor HSV (*Hue, Saturation and Value*) que define a cor de cada *pixel* baseado em matiz, saturação e valor de brilho. Baseado nesta composição tenta-se isolar a cor vermelha na imagem que é a parte correspondente ao final do contador de consumo localizado no medidor. Em seguida é aplicada a técnica de projeção horizontal e vertical na imagem com o objetivo de segmentar o contador de consumo e os dígitos em separado. A extração de características é feita baseada em comparações de descritores de perfil, cruzamento, contorno e histograma. Para evitar a ocorrência de vetores de características com tamanhos distintos devido à possibilidade de diferença de tamanho nos dígitos segmentados, todos os vetores gerados são forçados a ter o tamanho padronizado representando uma imagem de dimensão de 52x83 *pixels*. Possíveis espaços em branco provenientes de dígitos menores que esta dimensão recebem o valor zero em seu *pixel*.

O processo de extração de características é aplicado a um número aleatório de imagens resultando em uma matriz de características. Desta matriz uma porcentagem é tomada para treinar uma RNA e o restante para teste e validação. A RNA faz a classificação e o reconhecimento dos dígitos. Um conjunto de experimentos foi executado para avaliar a aplicação da RNA combinada com cada um dos descritores utilizados para a extração de características. Obteve-se uma base de 180 instâncias diferentes de cada dígito (0 até 9), totalizando 1800 elementos na base de imagens contendo os dígitos.

Dois cenários foram utilizados para validar os algoritmos, no primeiro selecionou-se aleatoriamente uma imagem com cada dígito em separado, e para cada dígito foi feita a tentativa de 100 reconhecimentos. O segundo cenário constou do reconhecimento de 100 instâncias de números selecionados de forma aleatória.

Os testes incluíram quatro RNAs com as unidades de entrada formadas por características extraídas do histograma dos dígitos. Tendo a camada intermediária configurada respectivamente com 5, 10, 15 e 20 neurônios, e dez saídas representando cada um dos dígitos (0 a 9). A configuração da RNA de 15 neurônios alcançou um nível de confiança de 90% e tempo abaixo de 6 milissegundos para cada dígito reconhecido.

Vanetti, Gallo e Nodari (2012) por sua vez utilizaram o algoritmo *MultiNOD* descrito no trabalho de Nodari e Gallo (2011) para o processo de segmentação, com o objetivo de investir contra a falta de robustez apresentada neste trabalho.

Para a detecção da região dos dígitos foi aplicado o algoritmo de *watershed* em uma etapa de pós-processamento da segmentação. Para as etapas de classificação e reconhecimento dos dígitos capturados foi utilizada a técnica de *Support Vector Machines* (SVM).

O trabalho de Belan, Araujo e Librantz (2013) propõe uma abordagem livre de segmentação para o reconhecimento de dígitos em instrumentos digitais de medição por meio de técnicas de casamento de padrões (*template matching*) em conjunto com correlação cruzada normalizada (NCC, do inglês *normalized cross-correlation*). No método descrito pelos autores, a imagem do instrumento é obtida por meio de uma *webcam*. Após transformação em tons de cinza o tamanho da imagem é reduzido em 30%. Esta redução do espaço de busca tem objetivo de acelerar o algoritmo de casamento de padrões que tem um custo alto de processamento. A técnica de casamento de padrões analisa uma imagem procurando por um determinado padrão pré-definido, neste caso utiliza padrões para os dígitos de 0 a 9. Como medida de similaridade utiliza a técnica de NCC que, de acordo com os autores, se mostra a forma mais robusta de medida de correlação para determinar a semelhança de pontos entre a imagem original e os padrões. Esta combinação de técnicas foi utilizada tanto para a determinação da região de interesse quanto para o reconhecimento dos dígitos. Para avaliar o desempenho do algoritmo proposto foram realizados testes com 50 imagens contendo diferentes tipos de ruídos, gerando o reconhecimento sem a apresentação de falsos positivos ou falsos negativos.

No trabalho de Edward (2013) é proposta uma solução para o reconhecimento dos dígitos de medidores de energia elétrica. Ao medidor é acoplada uma câmera disposta em uma posição que facilite a focalização e cobertura de todo o mostrador do medidor. Esta câmera é equipada com um temporizador que faz o papel de instruir a câmera na captura de imagens do medidor em intervalos regulares. Com a imagem capturada e transformada em tons de cinza, o autor aplica um filtro espacial linear para eliminação de possíveis ruídos provenientes do processo de captura da imagem, bem como propõe a normalização da imagem para padronizar sua altura e largura. A imagem é limiarizada, com o limiar definido utilizando a mediana de sua escala de cinza, a qual é calculada por meio da análise do histograma da imagem.

Para segmentar a área de interesse que contém os dígitos, o autor se baseia nas características de que a área contendo os dígitos possui um fundo preto contendo dígitos na cor branca e é uma região retangular de tamanho fixo. Para tal é utilizada a técnica de projeção vertical e horizontal dos *pixels* brancos da imagem. Isto feito, a imagem segmentada é normalizada para um tamanho padrão de 256x64 *pixels*.

Para detectar os dígitos é utilizada a técnica de janelas deslizantes de 64x64 *pixels* que percorrem a imagem para distinguir as janelas que possuem ocorrência de dígitos das que não possuem. Para esta distinção, foi utilizado um classificador baseado na técnica SVM treinado com uma base contendo 5000 exemplos de imagens de dimensões 64x64, considerando 2500 imagens com casos positivos para ocorrência de dígito e 2500 com casos negativos. Os casos que obtiveram retorno positivo para ocorrência de dígitos são apresentados a outro classificador SVM treinado apenas com os 2500 exemplos de casos positivos. O autor não destaca estatísticas numéricas de acerto no reconhecimento ou mesmo desempenho do algoritmo.

Em Zhao et al. (2013), a imagem processada não é relacionada diretamente a um medidor de energia, mas apenas individualmente ao ciclômetro, a peça que compõe o medidor de energia responsável por fazer a contagem do volume. Baseados nessa peça os autores capturam, com uma câmera comum, várias imagens contendo vários posicionamentos dos dígitos.

As imagens capturadas são transformadas em tons de cinza, e então a etapa de pré-processamento é iniciada com a aplicação da técnica de alongamento do contraste para melhorar a relação de contraste entre os dígitos e o fundo do ciclômetro. Na imagem resultante é aplicada uma operação morfológica *top-hat* para correção de ruídos provenientes de iluminação irregular. Na sequência, utilizam-se os métodos de projeção horizontal e vertical para encontrar os limites de altura e largura da região contendo os dígitos, desta forma a área de interesse na imagem é segmentada. Alguns ajustes ainda são efetuados como a eliminação de ruídos de espalhamento por meio da contagem de *pixels* conectados e o ajuste da inclinação dos dígitos contidos na região de interesse utilizando a análise da altura média da metade esquerda e direita da imagem.

Para segmentar os dígitos individualmente, primeiramente inverte-se a ocorrência de *pixels* pretos e brancos (dígitos passam a ser do grupo de *pixels* pretos). Percorre-se a imagem da base ao topo para encontrar a primeira ocorrência de um *pixel* preto, bem como do topo até a base na busca também da ocorrência do primeiro *pixel* preto, delimitando assim a altura aproximada da imagem. Percorre-se a imagem da esquerda para a direita para encontrar o primeiro *pixel* preto, considerando este o ponto inicial do dígito segmentado. Continua-se percorrendo em direção à direita até não encontrar mais *pixels* pretos, considerando este ponto como o final do dígito segmentado. Depois de segmentado o dígito individualmente, normaliza todos os caracteres para 20x30 *pixels*, utilizando interpolação bilinear.

Para a etapa de reconhecimento, o autor emprega uma Rede Neural utilizando para treinamento o algoritmo BP para reconhecimento dos caracteres, em uma base de treinamento de 100 exemplos de caracteres completos (0 até 9), gerando 10 classes no total e cada classe tendo 10 números. Similarmente, outra base de treinamento contendo 100 exemplos de dígitos incompletos é utilizada. Realiza-se o experimento de teste com 50 imagens contendo dígitos completos e incompletos obtendo 175ms de execução para uma taxa de 99,96% de acerto para os caracteres completos e 210ms de execução com taxa de 99,80% de acerto para os caracteres incompletos.

### 3.3 CONSIDERAÇÕES FINAIS

Nesta seção foram expostas as principais pesquisas que tratam do problema de reconhecimento ótico de dígitos apresentados em medidores de energia. Após a análise individual de cada trabalho, foi possível a identificação de variadas técnicas para a localização da ROI e segmentação individual dos dígitos, demonstrando que mesmo para objetos semelhantes, elaborar algoritmos aplicados ao problema de segmentação é uma das tarefas mais desafiadoras dentro da área de processamento de imagens e visão computacional conforme retrataram Gonzalez e Woods (2002). Da mesma forma foram expostas diversas técnicas relacionadas ao processo de extração de característica e reconhecimento dos dígitos apresentados no medidor de consumo.

De uma forma geral, os trabalhos buscaram segmentar a área contendo os dígitos partindo de uma imagem capturada ou transformada em tons de cinza. Neste trabalho é abordado um método de segmentação partindo da análise do espaço de cores da imagem.

Outro aspecto importante encontrado nos trabalhos é que, de uma forma geral, as imagens são capturadas em meio a um ambiente controlado utilizando câmeras fixas a uma determinada distância padrão do medidor de consumo, ou acopladas ao medidor, com a captura executada até mesmo em laboratório.

Neste trabalho a proposta é a de utilizar imagens de medidores capturadas em seu ambiente natural, no dia a dia dos profissionais que executam a tarefa de leitura do consumo.

Como forma de catalogar as diversas técnicas utilizadas nas variadas abordagens propostas pelos autores, é apresentada nos Quadros 2 e 3 uma visão global dos trabalhos encontrados na literatura reunindo as técnicas utilizadas em cada abordagem.

No Quadro 2, as técnicas são agrupadas com relação à tarefa de localização dos dígitos, as quais são técnicas relativas ao pré-processamento e segmentação, bem como técnicas relacionadas ao processo de OCR, que são relativas às etapas de extração de características e reconhecimento.



O quadro ainda destaca a taxa de acerto no processo de reconhecimento dos dígitos para o caso de trabalhos que destacaram este tipo de análise.

**Quadro 2. Resumo de técnicas relatadas na literatura.**

<b>Autor / Ano</b>	<b>Etapa</b>	<b>Técnicas Utilizadas</b>	<b>Taxa de Acerto</b>
<b>Alexandria et al. (2004)</b>	Captura	Câmera conectada a um microcomputador.	Não relatado
	Localização dos dígitos	- Limiarização. - Padronização da altura e largura dos dígitos segmentados (16x16).	
	Reconhecimento	- RNA (256 unidades de entrada, 64 neurônios intermediários e 10 saídas).	
<b>Castells-Rufas e Carrabina (2006)</b>	Captura	- Sensor CMOS + dispositivo FPGA acoplados ao medidor.	99,46%  base de 396 imagens
	Localização dos dígitos	- Limiarização pela média entre menor e maior valor de <i>pixels</i> .	
	Reconhecimento	- Máquina de estados finitos para extrair atributos. - Classificação por programação dinâmica (algoritmo de <i>Needleman-Wunsch</i> ).	
<b>Shu, Ma e Jing (2007)</b>	Captura	Câmera de vídeo ligada a uma placa de microcontroladora de captura, acoplados ao medidor.	Não relatado
	Localização dos dígitos	- Limiarização. - Projeção horizontal e vertical. - Operadores morfológicos de erosão e dilatação. - Inversão de cores.	
	Reconhecimento	- Extração de características baseadas nos traços de cada dígito individual (traços retos, curvos completos, curvos incompletos, posicionamento do traço no topo ou na base da imagem). - Projeção horizontal. - Rotulagem de componentes conexos.	
<b>Oliveira, Cruz e Bensebaa (2009)</b>	Captura	Câmera CCD.	Não relatado
	Localização dos dígitos	- Filtro homomórfico. - Transformação em tons de cinza.	

		<ul style="list-style-type: none"> <li>- Limiarização Otsu.</li> <li>- Operadores morfológicos de abertura, gradiente morfológico e fechamento.</li> <li>- Análise de componentes conexos.</li> </ul>	
	Reconhecimento	- Classificação por KNN ( <i>K-Nearest Neighbor</i> ).	
<b>Zhao et al. (2009)</b>	Captura	Sensor CMOS + microprocessador DSP.	95% base de 500 imagens
	Localização dos dígitos	<ul style="list-style-type: none"> <li>- Transformação em tons de cinza.</li> <li>- Equalização de histograma.</li> <li>- Projeção horizontal e vertical.</li> <li>- Padronização da altura e largura dos dígitos segmentados (21x15).</li> <li>- Limiarização.</li> </ul>	
	Reconhecimento	- RNA (315 unidades de entrada, 10 neurônios intermediários e 10 saídas).	
<b>Bai et al. (2010)</b>	Captura	Câmera acoplada ao medidor.	94% base de 50 imagens
	Localização dos dígitos	<ul style="list-style-type: none"> <li>- Transformação em tons de cinza.</li> <li>- Equalização de histograma.</li> <li>- Filtro da mediana.</li> <li>- Limiarização global.</li> <li>- Projeção horizontal e vertical.</li> </ul>	
	Reconhecimento	<ul style="list-style-type: none"> <li>- Momento invariantes de <i>Hu</i> para extrair atributos.</li> <li>- RNA (7 unidades de entrada, 14 neurônios intermediários e 10 saídas).</li> </ul>	

<b>Lei et al. (2010)</b>	Captura	Câmera em <i>handheld</i> PC.	99% base de 850 imagens
	Localização dos dígitos	<ul style="list-style-type: none"> <li>- Transformação em tons de cinza.</li> <li>- Filtro de média.</li> <li>- Limiarização (Otsu).</li> <li>- Operadores morfológicos de erosão e dilatação.</li> <li>- Interpolação bilinear.</li> <li>- Crescimento de regiões.</li> <li>- Esqueletização.</li> </ul>	
	Reconhecimento	- Classifica por meio de busca em vizinhança 8 no esqueleto dos dígitos.	
<b>Nava-Ortiz, Gómez e Diaz-Perez (2011)</b>	Captura	Câmera conectada a Micro PC (para treinamento) e câmera em <i>smartphone</i> (para testes).	93% base de 217 imagens
	Localização dos dígitos	<ul style="list-style-type: none"> <li>- Transformação em tons de cinza.</li> <li>- Limiarização (Bernsen).</li> <li>- Rotulagem de componentes conexos.</li> <li>- Padronização da altura e largura dos dígitos segmentados (20x20).</li> </ul>	
	Reconhecimento	Classificadores de distância mínima Euclidiana e Mahalanobis.	
<b>Nodari e Gallo (2011)</b>	Captura	Não relatada.	58%
	Localização dos dígitos	- Árvore de RNAs (denominada <i>MultiNOD</i> ).	
	Reconhecimento	- OCR Tesseract.	
<b>Jabba et al. (2012)</b>	Captura	Câmera wireless conectada a micro PC.	
	Localização dos dígitos	- Análise do espaço de cor HSV.	

		- Projeção horizontal e vertical.	90% base de 1800 imagens
	Reconhecimento	- Descritores estatísticos de perfil, cruzamento, contorno e histograma para extrair atributos.  - Padronização da altura e largura dos dígitos segmentados (52x83).  - RNA (15 unidades de entrada e 10 de saída).	
<b>Vanetti, Gallo e Nodari (2012)</b>	Captura	Não relatada.	Não relatado
	Localização dos dígitos	- Árvore de RNAs (denominada <i>MultiNOD</i> ).  - <i>Watersheds</i> .	
	Reconhecimento	- <i>Support Vector Machines (SVM)</i> .	
<b>Belan, Araujo e Librantz (2013)</b>	Captura	<i>Webcam</i> conectada a micro PC.	Não relatado
	Localização dos dígitos e Reconhecimento	- Transformação em tons de cinza.  - <i>Template matching</i> .  - Correlação cruzada normalizada.	
<b>Edward (2013)</b>	Captura	Câmera acoplada ao medidor.	Não relatado
	Localização dos dígitos	- Transformação em tons de cinza. - Filtro espacial linear. - Projeção horizontal e vertical. - Padronização da altura e largura dos dígitos segmentados (64x64). - Janelas deslizantes.	
	Reconhecimento	- <i>Support Vector Machines (SVM)</i> .	

<b>Zhao et al. (2013)</b>	Captura	Câmera comum.	99%  base de 50 imagens
	Localização dos dígitos	- Transformação em tons de cinza. - Alongamento de contraste. - Operador morfológico <i>top-hat</i> . - Projeção horizontal e vertical. - Padronização da altura e largura dos dígitos segmentados (20x30).	
	Reconhecimento	- RNA.	

O Quadro 3 contabiliza a distribuição das técnicas utilizadas nos trabalhos encontrados na literatura. A última coluna indica o número de vezes que cada técnica foi utilizada dentre os trabalhos analisados.

**Quadro 3. Contabilização das técnicas relatadas na literatura.**

Técnica utilizada	Autor / Ano											T o r a l			
	Alexandria et al. (2004)	Castells-Rufas e Carrabina (2006)	Shu, Ma e Jing (2007)	Oliveira, Cruz e Bensebaa (2009)	Zhao et al. (2009)	Bai et al. (2010)	Lei et al. (2010)	Nava-Ortiz, Gómez e Diaz-Perez (2011)	Nodari e Gallo (2011)	Jabba et al. (2012)	Vanetti, Gallo e Nodari (2012)		Belan, Araujo e Librantz (2013)	Edward (2013)	Zhao et al. (2013)
Análise do espaço HSV										X					1
<i>Contrast Stretching</i>														X	1
Correlação cruzada normalizada												X			1



Operadores morfológicos de dilatação			X				X											2	
Operadores morfológicos de erosão			X				X												2
Operadores morfológicos de fechamento				X															1
Operadores morfológicos de gradiente				X															1
Padronização da altura e largura dos dígitos segmentados	X				X			X		X				X		X			6
Programação dinâmica por <i>Needleman-Wunsch</i>		X																	1
Projeção horizontal			X		X	X				X				X		X			6
Projeção vertical			X		X	X				X				X		X			6
Rede neural artificial	X				X	X			X	X	X					X			7
Rotulagem de componentes conexos			X	X				X											3
<i>Support Vector Machine (SVM)</i>											X			X					2
Casamento de padrões												X							1
Transformação em tons de cinza				X	X	X	X	X					X	X	X				8
<i>Watersheds</i>											X								1

Os trabalhos encontrados na literatura serviram como base para a elaboração do método proposto nesta dissertação, o qual será descrito próxima seção.

## 4 MATERIAIS E MÉTODOS

### 4.1 INTRODUÇÃO

O presente capítulo descreve as principais características dos materiais e métodos empregados no desenvolvimento desta dissertação. A metodologia empregada nesta pesquisa teve como objetivo a obtenção de um algoritmo que processe imagens contendo um medidor de consumo de gás natural, localize o contador que apresenta o volume de gás consumido, identifique os dígitos significativos e faça o reconhecimento e apresentação destes em formato texto.

Entenda-se o termo “dígitos significativos” como dígitos pertencentes ao número apresentado na parte preta do contador de consumo, pois, por convenção das empresas distribuidoras de gás natural a leitura do consumo é feita descartando-se os dígitos apresentados na parte vermelha do contador.

Nos itens subsequentes serão descritas as características relacionadas a materiais utilizados nesta pesquisa, bem como os passos que compõem o método proposto para a construção do algoritmo e obtenção dos resultados desejados.

### 4.2 MATERIAIS

Neste trabalho, o algoritmo foi projetado para lidar com imagens de medidores que possuam as seguintes características:

- Medidores com 4, 5, 6 ou 7 dígitos significativos (dígitos de 0 a 9);
- Dígitos significativos em cor clara sobre fundo escuro;
- Dígitos descartáveis em cor clara sobre fundo vermelho ou contidos em um retângulo englobante de bordas vermelhas;
- Imagens capturadas do cotidiano conforme retrata a Figura 22, representando situações reais encontradas por profissionais que efetuam mensalmente a



leitura, porém sem apresentar obstruções que sejam capazes de impossibilitar o reconhecimento dos dígitos.



Figura 22. Imagens de medidores em situações do cotidiano.

Foram utilizadas 793 imagens de medidores capturadas em situações diversas, apresentando variedade quanto aos aspectos de aproximação, iluminação, e nível de ruído, e em alguns casos inclinação.

As imagens fazem parte de dois grupos de imagens. O primeiro grupo (G1) originado de imagens obtidas junto ao setor responsável pela leitura do consumo mensal de gás dos clientes da Companhia Paranaense de Gás (COMPAGAS). O segundo grupo (G2), oriundo de imagens disponibilizadas pelo laboratório de pesquisas italiano Arte-Lab<sup>1</sup>, mesma base utilizada em um dos trabalhos encontrados na literatura (NODARI; GALLO, 2011). Vale ressaltar que no trabalho do autor a segmentação da ROI não leva em consideração o espaço de cores da imagem, partindo de uma imagem em tons de cinza para obter a segmentação. Desta forma, em alguns casos, imagens do grupo G2 foram descartadas por terem sido capturadas sem destacar de forma suficiente a parte vermelha do contador de consumo, fato que inviabilizaria o método de segmentação da ROI aqui proposto.

Para o experimento foram utilizadas 640 imagens de medidores do grupo G1 originando 930 dígitos individuais e 153 imagens de medidores do grupo G2 originando 290 dígitos individuais. O número relativamente baixo de dígitos, se comparado ao número de imagens de medidores, se deve ao fato de que a segmentação individual gerou quantidades diferentes para cada dígito (0 a 9). Por

<sup>1</sup> Banco de imagens disponível em: <http://artelab.dicom.uninsubria.it/downloads.html>, ver item "Gas-meter reading" (acessado em 01/06/2016).

exemplo, o dígito zero apresentou 1790 ocorrências enquanto que o dígito nove apresentou 122 ocorrências. Desta forma, para que a base não ficasse desbalanceada foram selecionadas aleatoriamente 122 ocorrências para cada um dos dígitos de 0 a 9, totalizando as 1220 amostras de dígitos obtidos utilizando o método aplicado para segmentação individual dos dígitos. Este método é apresentado nas próximas seções.

### 4.3 MÉTODOS

Analisando a distribuição das técnicas empregadas na literatura, observamos diversas abordagens cujo trabalho de processamento da imagem é aplicado a partir de imagens transformadas ou capturadas em tons de cinza (13 dentre os 14 trabalhos levantados no estado da arte).

Em razão disto, o método aqui proposto envolve a segmentação levando-se em conta o espaço de cores da imagem, visto que uma característica importante observada nas imagens de medidores é o padrão encontrado na área do contador. Este é formado por dígitos brancos aplicados sobre uma área preta conectada a uma área vermelha, conforme apontado na Figura 23.



**Figura 23. Área vermelha apresentada no contador de consumo dos medidores.**

A principal dificuldade encontrada no decorrer do desenvolvimento do método proposto neste trabalho teve relação com a tarefa de segmentação da ROI e dos dígitos de forma individual.

Os trabalhos levantados no estado da arte apresentam soluções diversas que de alguma forma contribuíram para o amadurecimento do método proposto. No

entanto, a maioria dos trabalhos empregam soluções que envolvem ambientes totalmente ou parcialmente controlados, o que inviabiliza a aplicação de determinadas técnicas. A proposta desta pesquisa foi a de empregar o método sugerido em imagens obtidas a partir de condições mais próximas do ambiente real, vivenciado no dia a dia dos profissionais que realizam a leitura do valor de consumo nos medidores.

No método empregado nesta pesquisa foram utilizadas variadas técnicas que compreendem a transformação do espaço de cores da imagem, limiarização, análise de projeções de *pixels*, detecção de linhas, morfologia matemática, rotulagem de componentes conexos, redimensionamento da imagem, extração de características da imagem, reconhecimento de características e interpretação. A Figura 24 mostra uma visão geral dos passos e principais técnicas utilizadas no método proposto.

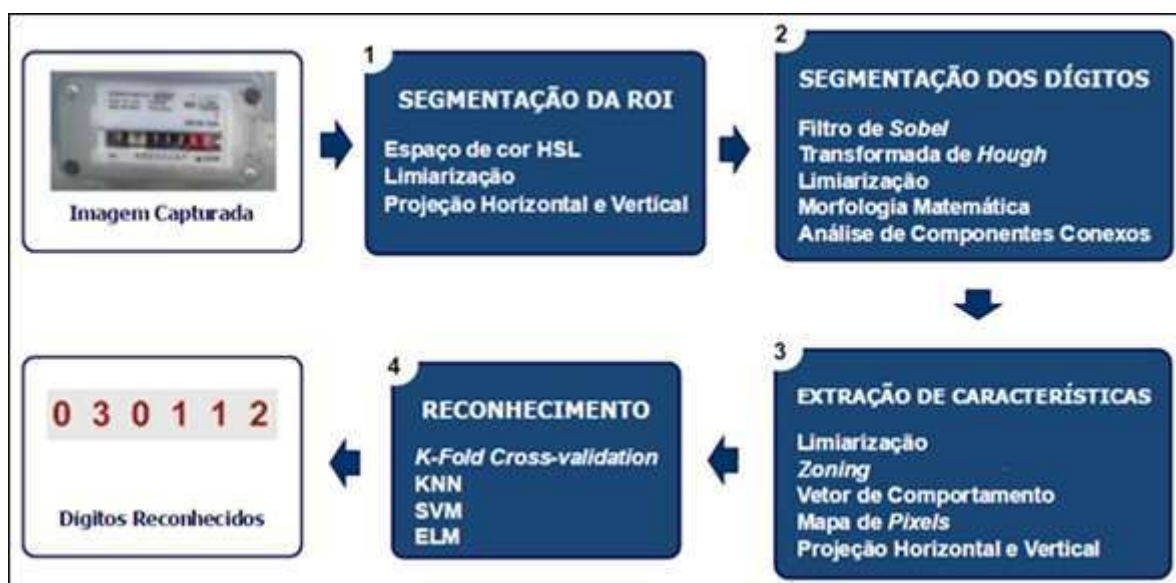


Figura 24. Resumo dos passos e principais técnicas utilizadas no método.

As próximas seções descrevem detalhadamente cada uma das fases do método proposto para a construção do algoritmo e obtenção dos resultados desejados.

#### 4.3.1 Segmentação da área contendo o contador de consumo

O objetivo desta etapa é o de localizar e segmentar a região que apresenta o contador de consumo que contém os dígitos a serem reconhecidos. Os passos e

técnicas utilizadas nesta etapa são ilustrados na Figura 25. Na sequência, são descritas cada uma das técnicas empregadas nesta etapa.



Figura 25. Passos para segmentação da ROI.

#### 4.3.1.1 Conversão do espaço de cores RGB para HSL

A abordagem proposta trata do uso do espaço de cor. Com a conversão da distribuição de valores RGB da imagem para a distribuição de valores HSL busca-se gerar uma imagem binária (limiarizada), na qual sejam observadas porções que indiquem a presença de *pixels* vermelhos (ou próximos de vermelho). Bem como uma imagem limiarizada com porções que indiquem a presença de *pixels* pretos (ou próximos de preto).

Para o caso de *pixels* vermelhos a análise é feita nos três canais H, S e L. Já no caso de *pixels* pretos apenas o canal L é analisado.

Baseado em experimentos aplicados à base de imagens foram definidos os seguintes limiares para a filtragem do espaço HSL:

- $H < 15$  ou  $H > 330$ ,  $S > 20$  e  $L > 10$ , para indicar a presença de pixels vermelhos.
- $L < 33$  para indicar a presença de pixels pretos.

A Figura 26 exibe o resultado da aplicação da limiarização com base nos limiares anteriormente citados.

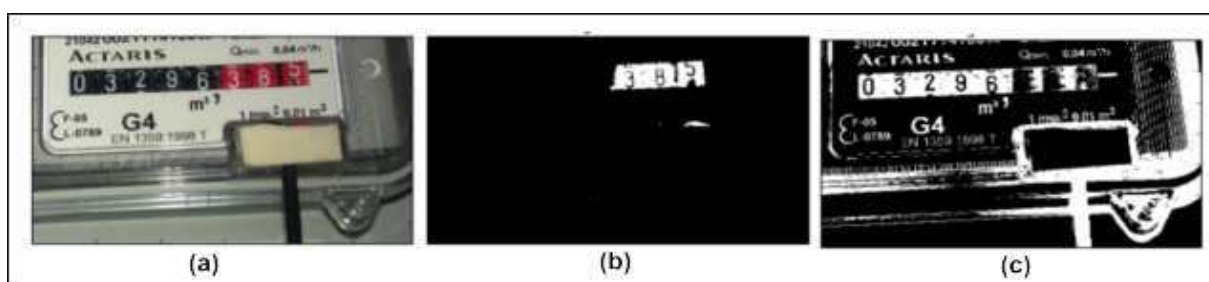


Figura 26. Limiarização de *pixels* vermelhos e *pixels* pretos. (a) Imagem original; (b) Limiarização – vermelhos; (c) Limiarização – pretos.

#### 4.3.1.2 Análise da projeção horizontal e vertical dos *pixels*

Após obter as imagens com os possíveis *pixels* vermelhos, gera-se para cada imagem a respectiva projeção horizontal (soma de *pixels* por linha). Este processo visa auxiliar a identificação de limiares para a altura preliminar da região de interesse em cada imagem.

A fim de atenuar as variações que ocorrem na projeção horizontal aplica-se um filtro com o objetivo de suavizá-las. O filtro analisa sucessivamente cada valor apresentado no vetor de projeções.

O processo de suavização utiliza uma máscara de 21 posições cujo elemento central é posicionado sobre cada elemento do vetor. O novo valor a ser atribuído a cada elemento do vetor é baseado no cálculo da média dos valores posicionados dez elementos à direita e dez elementos à esquerda do elemento em análise.

Na Figura 27 pode ser observada a forma como o filtro de suavização é aplicado em determinado elemento do vetor de projeções.

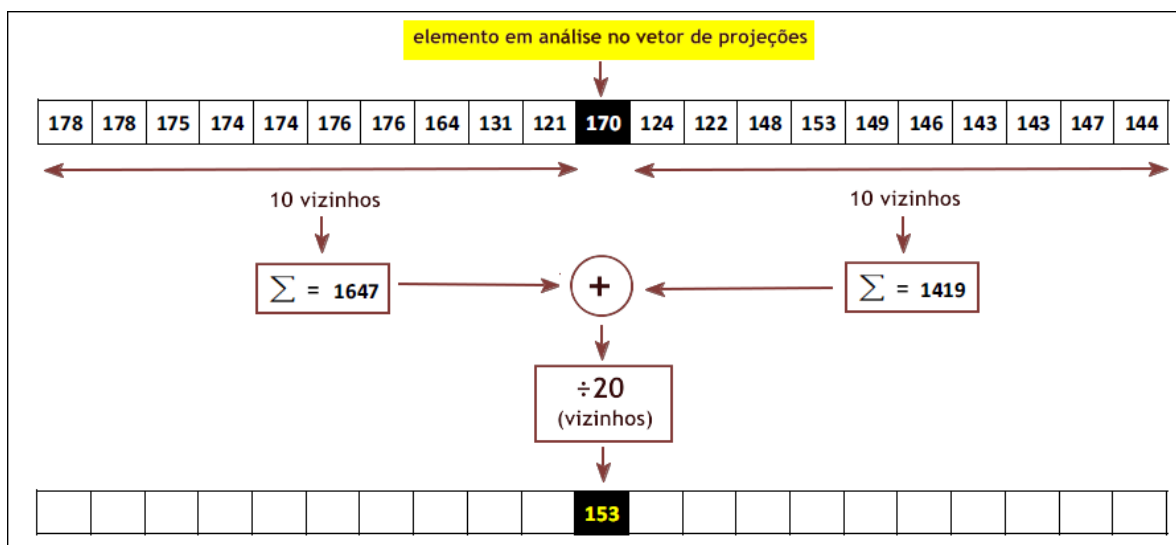


Figura 27. Aplicação do filtro de suavização em elemento do vetor de projeções.

O objetivo da suavização é atenuar a variação de picos e vales apresentada pelo histograma de projeções e facilitar a localização de picos e vales na projeção. A dimensão de 21 posições para a máscara de suavização foi definida baseada em experimentos aplicados à base de imagens.

A Figura 28 apresenta os gráficos resultantes da projeção horizontal e da projeção horizontal suavizada para imagem apresentada na Figura 26(b).

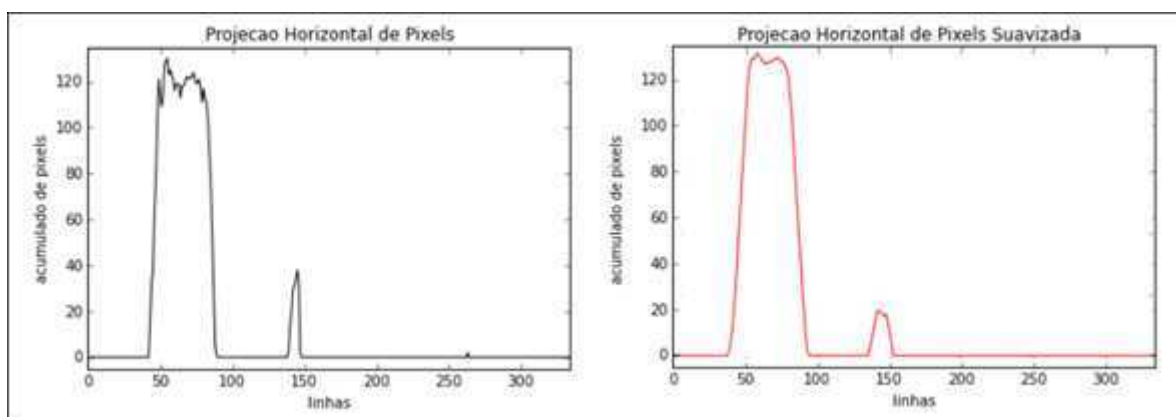


Figura 28. Projeções horizontais de *pixels*. (a) Original; (b) Suavizada.

#### 4.3.1.3 Análise de picos e vales

Com o objetivo de identificar os picos e vales presentes na projeção suavizada, aplica-se um cálculo levando-se em conta uma vizinhança de três elementos à direita e três à esquerda do elemento analisado no vetor de projeções.

O novo valor do elemento é obtido por meio da subtração do valor encontrado no seu terceiro vizinho mais à direita pelo valor do seu terceiro vizinho mais à esquerda. Os valores que resultam positivos identificam os picos no vetor de projeções, e os negativos identificam os vales.

A quantidade de três elementos levada em consideração no cálculo, foi obtida mediante experimentos aplicados à base de imagens. A Figura 29 ilustra o modo de cálculo utilizado para encontrar os picos e vales no vetor de projeções suavizado.

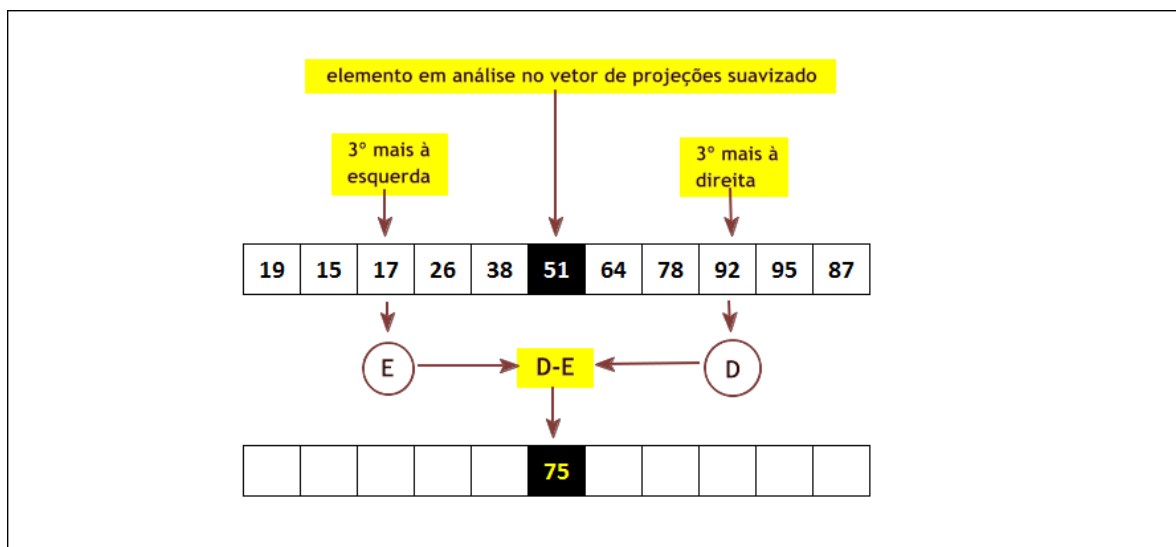


Figura 29. Forma de localização de picos e vales no vetor de projeções.

A Figura 30 apresenta os gráficos resultantes da identificação de picos e vales para a projeção horizontal suavizada apresentada na Figura 28(b).

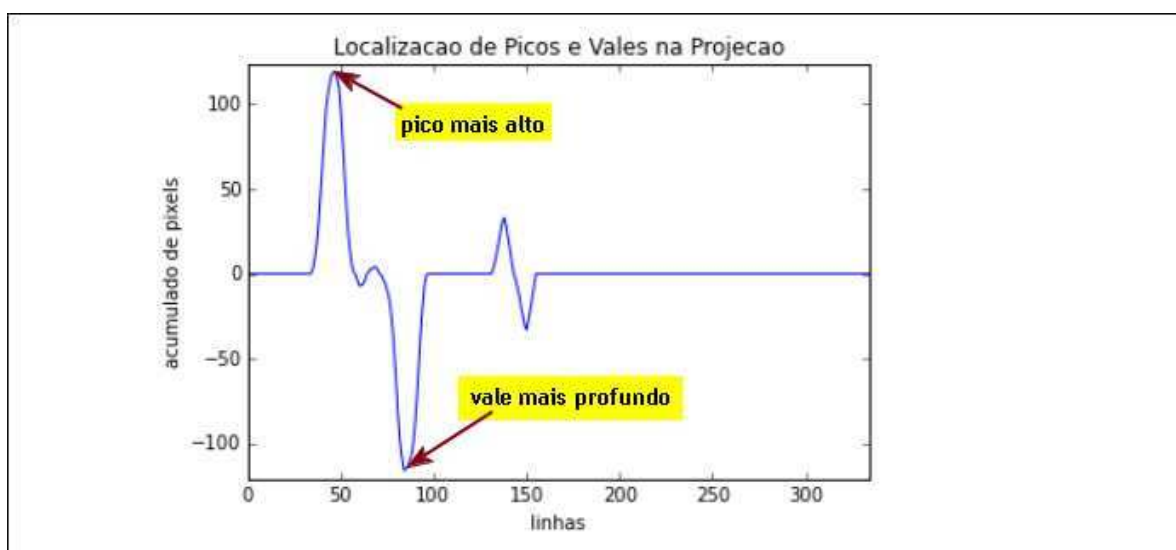


Figura 30. Identificação de picos e vales na projeção suavizada.

A combinação da identificação do pico mais alto com o vale mais profundo encontrado à sua frente irá delimitar de forma preliminar a altura da ROI (a altura é determinada pelo pico que identificará a linha inicial da ROI e o vale que identificará a linha final da ROI). Se a altura encontrada for menor que uma altura mínima, busca-se a combinação com o próximo vale mais profundo até que a altura mínima da ROI seja atingida. Baseado em experimentos aplicados à base de imagens, o limiar de altura mínima foi definido em 25 *pixels*.

Encontrado os limites inicial e final para a altura preliminar da ROI, aplica-se a estes uma tolerância de 30 *pixels* acima do limite inicial e abaixo do limite final. Isto se fez necessário para evitar segmentações abruptas relacionadas a imagens que apresentem um aspecto de inclinação, permitindo assim que o contador de consumo mesmo com certo grau de inclinação seja apresentado de forma completa na segmentação preliminar da ROI.

#### 4.3.1.4 Segmentação da região de interesse na largura

Tendo identificado a linha inicial e final que delimitam de forma preliminar a altura da ROI, tomam-se as imagens resultantes do processo anterior de limiarização. A partir das imagens, uma que identificou regiões vermelhas e outra que identificou regiões pretas, utiliza-se os limites preliminares obtidos para segmentá-las em sua altura, conforme apresentado na Figura 31.



**Figura 31. Segmentação do resultado da limiarização na altura. (a) Limiarização – vermelhos; (b) Limiarização – pretos.**

A partir da imagem representando a área vermelha da ROI, percorre-se esta da esquerda para a direita analisando-se a projeção vertical (soma dos *pixels* por coluna) até encontrar o limite inicial da região vermelha, o que delimita a coluna final da ROI (final da região preta contendo o contador de consumo).

Porém, observou-se em algumas imagens a presença de ruídos apresentados em colunas anteriores ao limite inicial da região vermelha. Desta forma, é aplicado



sobre a imagem limiarizada de *pixels* vermelhos, um operador morfológico de erosão com elemento estruturante quadrado de dimensão três. A aplicação deste operador tem como objetivo evitar que um ruído anterior ao limite procurado possa ser definido como um falso limite inicial da região vermelha, e por consequência a identificação incorreta da coluna final da ROI.

Tendo identificado o limite da coluna final da ROI, analisa-se a imagem representando a área preta, tomando como início o limite identificado para a coluna final da ROI. Percorre-se então da direita para a esquerda analisando-se a projeção vertical de cada coluna até encontrar o limite no qual se identifica o início do contador de consumo, que por consequência determina a coluna inicial da ROI. O processo, que delimita a largura da ROI, é representado na Figura 32.

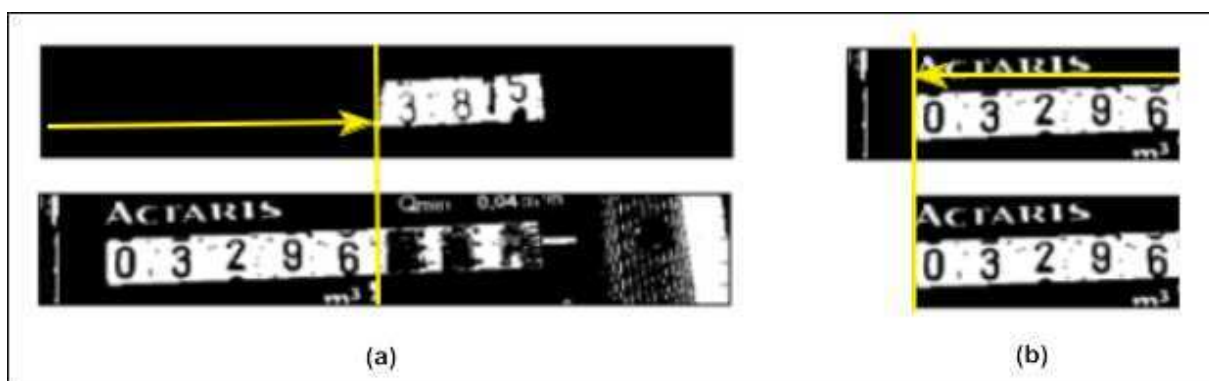


Figura 32. Delimitação da largura da ROI. (a) Coluna final localizada pela análise da área vermelha e projetada na área preta; (b) Coluna inicial localizada pela análise da área preta.

#### 4.3.1.5 Ajuste da inclinação

Para segmentar a ROI em sua altura final, é necessário antes verificar se o resultado obtido até o momento não apresenta uma imagem com inclinação. Pois se isto ocorrer, a sequência do método que segmentará a ROI em sua altura final poderá sofrer interferência trazendo resultados indesejados, como por exemplo, cortar parte do início ou final de algum dígito.

A abordagem para verificar se a imagem apresenta inclinação, e em caso afirmativo ajustá-la, baseia-se no cálculo do ângulo de inclinação das bordas horizontais da imagem e para tanto se cumprem os seguintes passos:

1. Segmentar a imagem original com base nos já identificados limites preliminares de altura e largura;

2. Transformar o resultado em tons de cinza;
3. Executar a limiarização do resultado utilizando Otsu;
4. Detectar bordas horizontais utilizando Sobel;
5. Detectar o ângulo de inclinação utilizando a transformada de Hough;
6. Rotacionar a imagem com base no ângulo de inclinação detectado.

A Figura 33 mostra as imagens resultantes do método para correção da inclinação da ROI preliminarmente detectada.

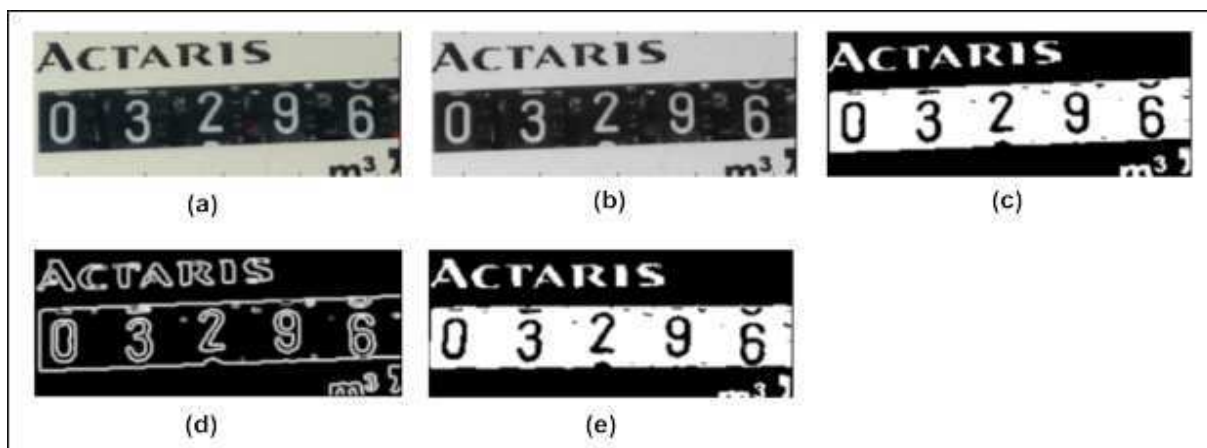


Figura 33. Ajuste de inclinação da ROI. (a) Imagem original segmentada; (b) Transformação em tons de cinza; (c) Limiarização; (d) Detecção de bordas horizontais; (e) Ajuste da inclinação.

#### 4.3.1.6 Segmentação da região de interesse na altura

Com a inclinação ajustada, o próximo passo é o de segmentar a altura definitiva da ROI. Para tanto, calcula-se a projeção horizontal da imagem e com base na projeção de cada linha, partindo da linha central da imagem, busca-se localizar a primeira ocorrência de linha que possua ao menos 70% dos seus *pixels* formados por *pixels* de fundo, tanto em direção ao topo para identificar o limite inicial de altura da ROI, quanto em direção à base para localizar o limite de altura final da ROI. A Figura 34 retrata a localização destes limites.

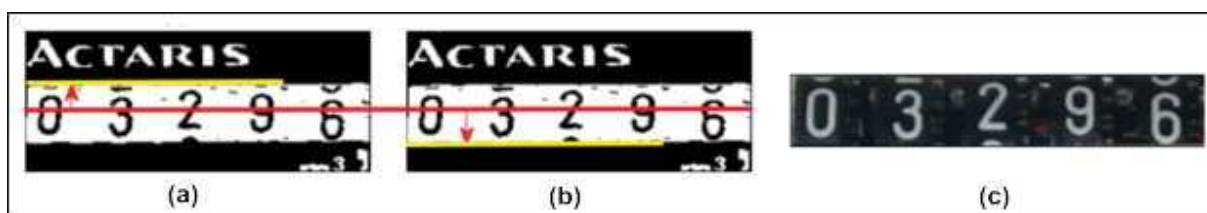


Figura 34. Segmentação da altura definitiva da ROI. (a) Limite de altura inicial; (b) Limite de altura final; (c) Segmentação definitiva.

### 4.3.2 Segmentação dos dígitos

Esta etapa tem como propósito a localização dos dígitos apresentados na ROI anteriormente segmentada, bem como a segmentação de cada dígito em imagens individuais. A Figura 35 apresenta os passos principais da abordagem empregada nesta etapa. Em seguida, são descritas as técnicas que fazem parte da abordagem.

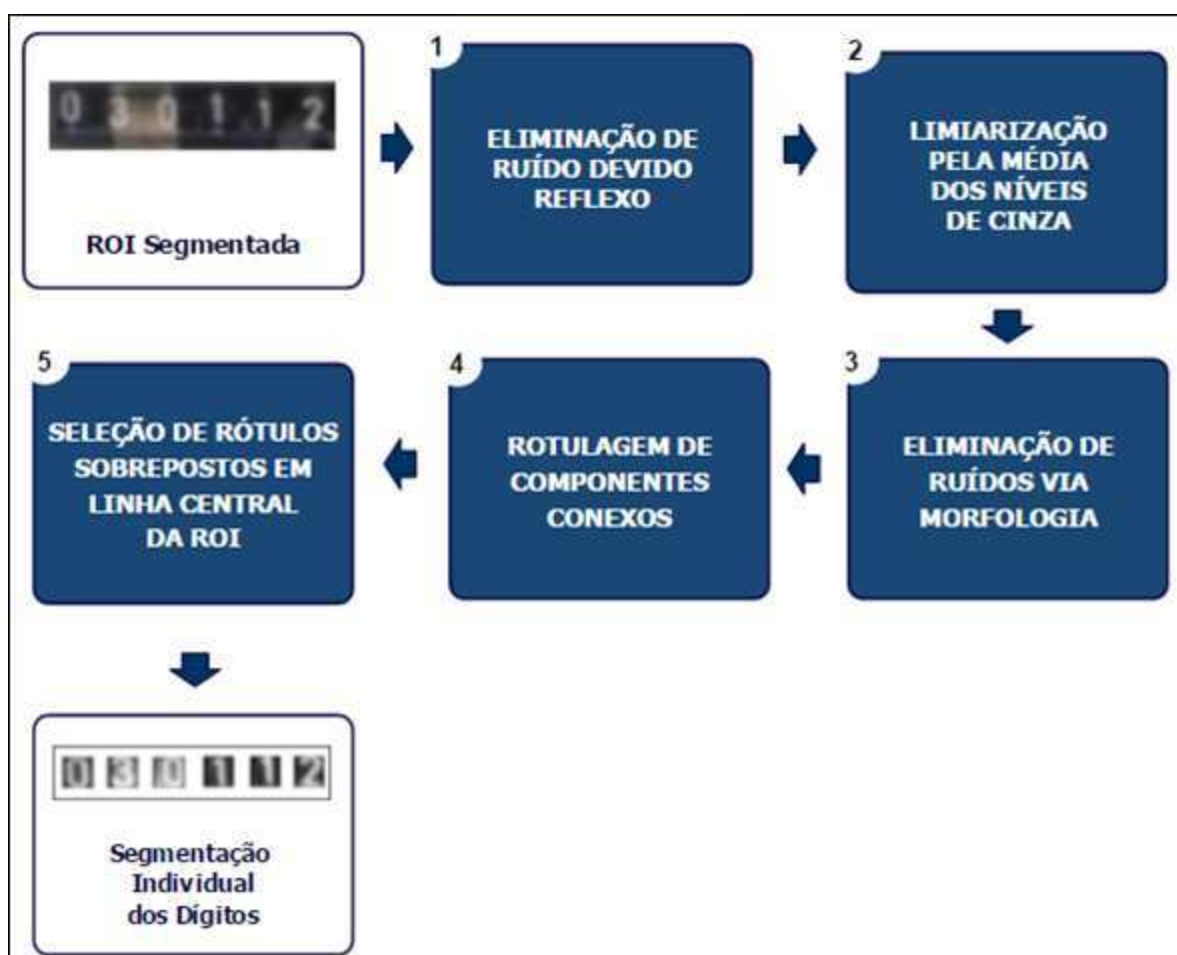
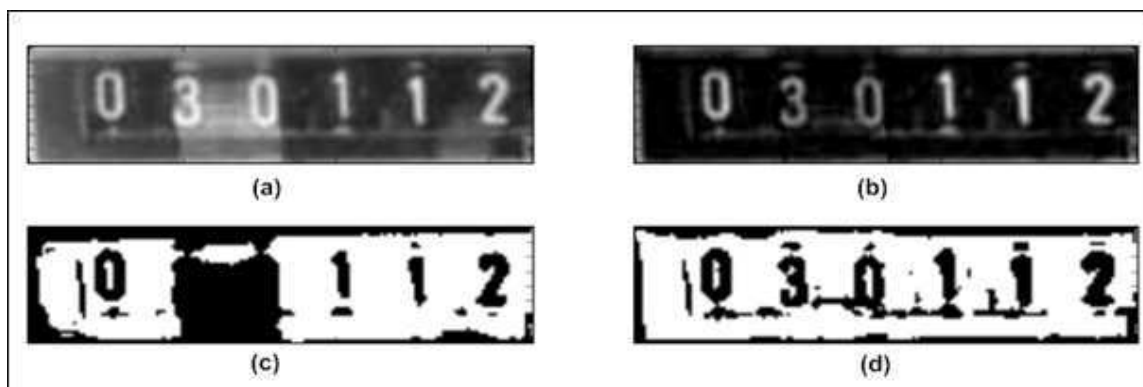


Figura 35. Passos para segmentação dos dígitos.

Para a segmentação individual dos dígitos aplica-se a técnica de Análise de Componentes Conexos (ACC), porém antes da aplicação da ACC algumas operações morfológicas são aplicadas com o objetivo de obter um melhor resultado na segmentação individual dos dígitos.

O operador morfológico *top-hat* é aplicado na ROI segmentada com o objetivo de eliminar ruídos provenientes do reflexo apresentado no vidro que cobre a ROI e assim auxiliar a tarefa de limiarização. Sem a adoção deste operador, se tornaria

praticamente impossível segmentar determinados dígitos como é o caso dos dígitos apresentados na Figura 36.



**Figura 36. Filtro *top-hat*. (a) ROI segmentada; (b) ROI após filtro *top-hat* (c) Limiarização sem *top-hat*; (d) Limiarização após *top-hat*.**

Após a aplicação da operação *top-hat* é aplicada a limiarização na ROI cujo limiar é definido como o valor médio dos níveis de cinza apresentados na ROI. Como o resultado ainda apresenta ruídos, conforme pode ser verificado na Figura 36(d), são aplicados na sequência dois outros operadores morfológicos.

Primeiramente, é aplicado o operador de fechamento com elemento estruturante retangular horizontal de tamanho 1x90. Em seguida a imagem resultante é invertida, conforme ilustra a Figura 37(b), e somada com a imagem ruidosa. Este procedimento é realizado, pois em alguns casos, após a limiarização, ocorre a presença de dígitos conectados com as linhas de borda do retângulo englobante que envolve o espaço contendo os dígitos conforme apresentado na Figura 37(a). A aplicação do operador de fechamento destaca as linhas de borda do retângulo englobante. A soma da imagem ruidosa com a imagem resultante do fechamento procura eliminar toda ou parte destas linhas da imagem ruidosa, e por consequência desconecta da borda do retângulo englobante os *pixels* que pertencem ao dígito.

A fim de eliminar ruídos que entremeiam os dígitos, é aplicado o operador de dilatação com elemento quadrado de tamanho 3x3. Um exemplo do resultado da aplicação destes dois operadores é apresentado na Figura 37.

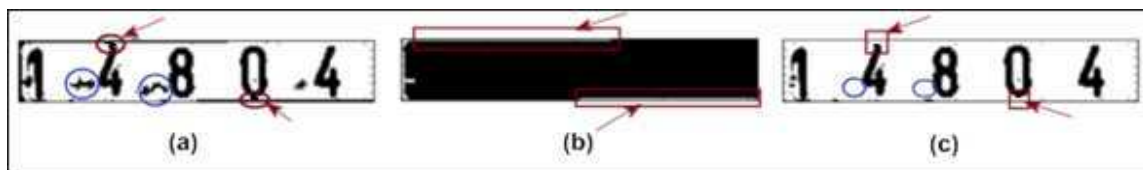


Figura 37. Eliminação de ruídos. (a) Imagem ruidosa; (b) Filtro morfológico de fechamento com resultado invertido; (c) Ruídos eliminados.

Baseado na premissa de que parte dos *pixels* que pertencem aos dígitos compõe a linha central da ROI, analisa-se a área de um segmento central composto de dez linhas abaixo e dez linhas acima a partir da metade da altura da ROI.

Desta forma serão considerados para análise apenas os rótulos resultantes da ACC que tenham *pixels* pertencentes a este segmento central, descartando-se os demais rótulos.

A partir dos rótulos selecionados, consideram-se como candidatos a dígitos apenas os rótulos que estiverem dentro de um determinado limite mínimo e máximo de *pixels*. Baseado em experimentos aplicados à base de imagens, o limiar para o mínimo foi definido em 10 *pixels* e para o máximo em 70 *pixels*, a Figura 38 retrata o resultado do procedimento descrito acima.

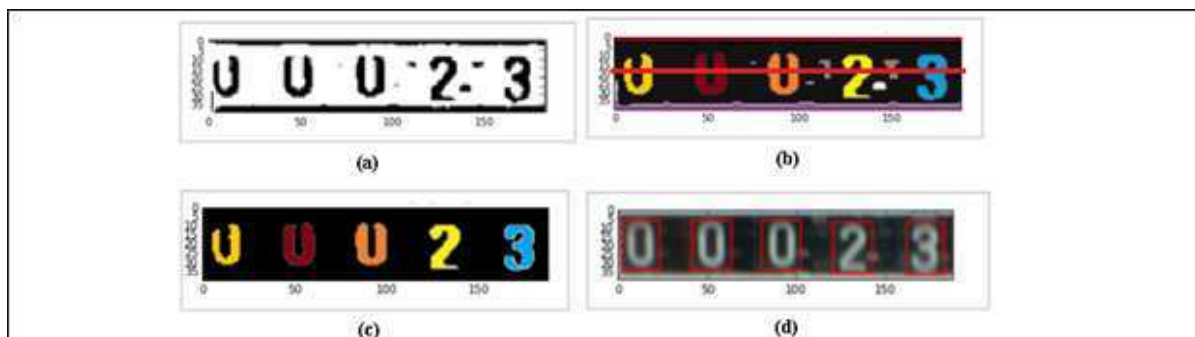
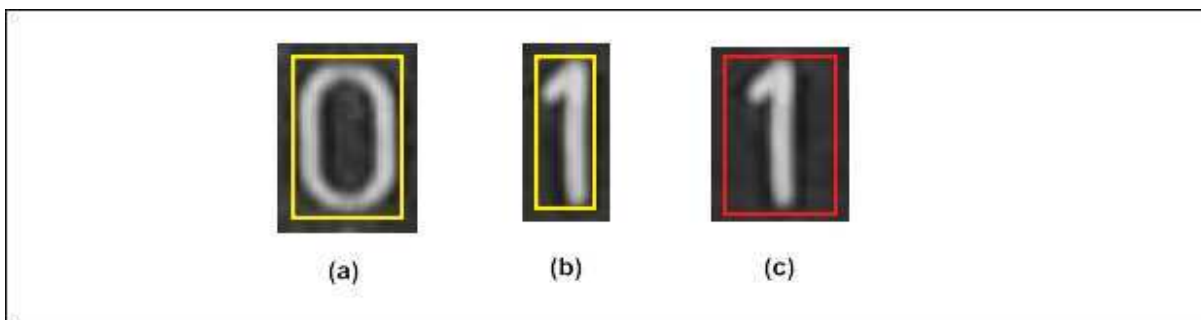


Figura 38. Segmentação dos dígitos. (a) Resultado da limiarização; (b) Resultado da rotulagem e linha central de análise; (c) Rótulos selecionados; (d) Dígitos segmentados.

Identificados os rótulos que representam os dígitos, é feito um ajuste para que todos os rótulos apresentem suas coordenadas com a mesma dimensão de altura e largura do rótulo com maior valor de altura e largura. Por exemplo, para o caso do dígito 1, o valor de distância entre os pontos que compõem a sua coordenada de largura deverá ser o mesmo valor de distância apresentado entre os pontos que compõe as coordenadas que definem a largura do dígito 0. A Figura 39 apresenta este tipo de ajuste.



**Figura 39. Ajuste pelo rótulo de dimensão máxima. (a) Rotulagem com maior dimensão de altura x largura; (b) Rotulagem com dimensão inferior; (c) Rotulagem com dimensão ajustada.**

Além do ajuste de coordenadas, é feito novo ajuste para que os dígitos segmentados de forma individual passem a apresentar um tamanho padrão. Neste trabalho foi definido que os dígitos terão um padrão de 40 *pixels* de largura por 40 *pixels* de altura.

Estes ajustes são necessários a fim de colaborar com o desempenho da próxima fase da metodologia que trata do reconhecimento dos dígitos, cujas técnicas utilizadas são descritas na próxima seção.

#### 4.3.6 Reconhecimento dos dígitos

Com o objetivo de projetar um classificador para realizar a tarefa de reconhecer os dígitos segmentados, foram definidas fases que envolveram:

- A criação de base de conhecimento contendo imagens de dígito de 0 a 9.
- A definição de atributos que caracterizasse cada dígito.
- A aplicação do classificador que retornasse o reconhecimento do dígito.

A Figura 40 resume os passos executados para a geração de atributos e classificação dos dígitos. Na sequência, é descrito o método utilizado para a geração da base de conhecimento, definição de atributos e reconhecimento dos dígitos.

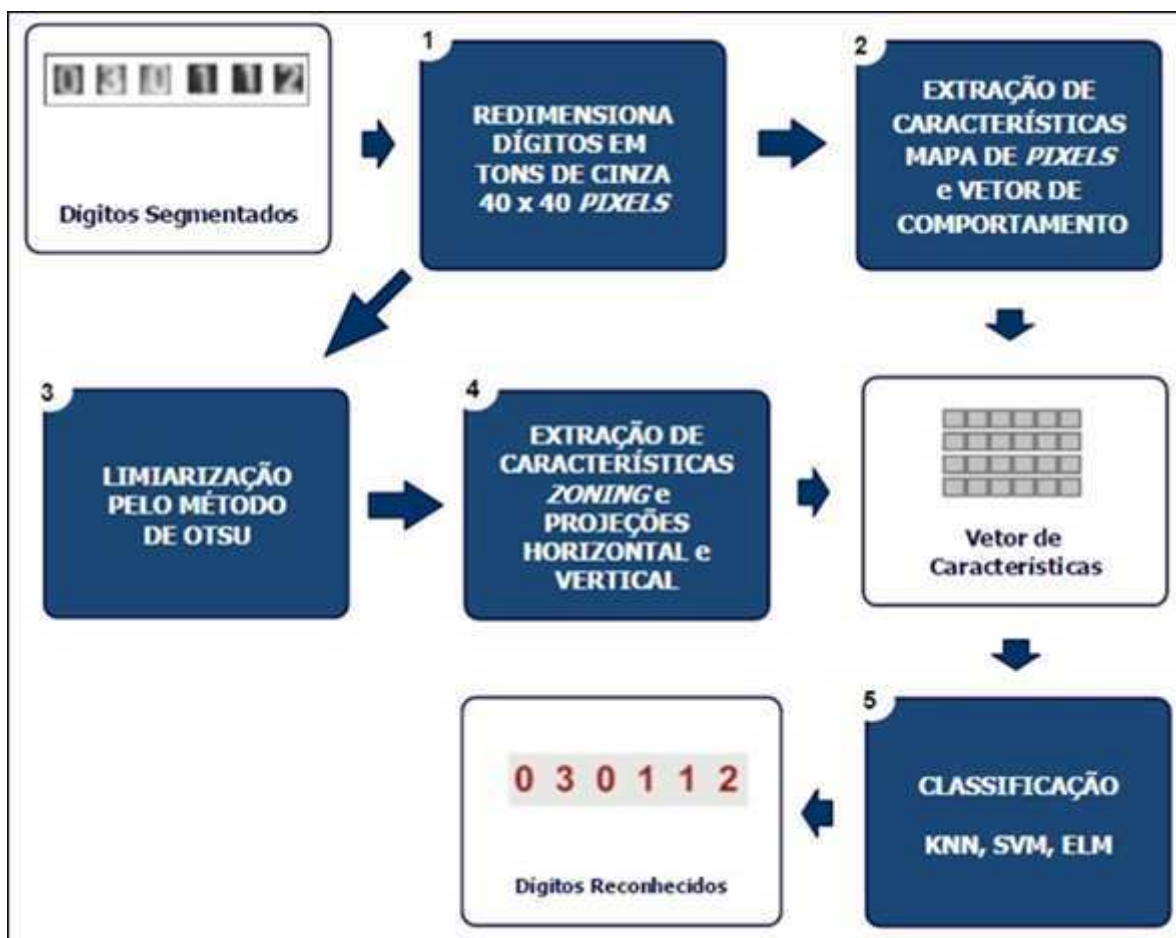


Figura 40. Resumos dos passos para o reconhecimento dos dígitos.

#### 4.3.6.1 Construção da base de amostras de dígitos

Utilizando as imagens de dígitos obtidos por meio do método de segmentação descrito neste trabalho, foi criada uma base de conhecimento que serviu para o treinamento e teste do método de classificação e reconhecimento individual dos dígitos.

Esta base foi construída utilizando-se uma aplicação desenvolvida com o propósito de atribuir um rótulo a cada um dos dígitos resultantes do processo de segmentação, conforme ilustrado na Figura 41.



Figura 41. Tela da aplicação que atribui rótulos aos dígitos segmentados.

Na aplicação, de forma manual, é atribuída para cada imagem apresentada em tela uma classificação correspondente ao valor identificado visualmente para o dígito segmentado. Estes valores, correspondentes a classificação manual de cada dígito, são armazenados em um arquivo texto contendo o rótulo de cada dígito. Este arquivo texto é a fonte de dados para a geração dos vetores de características obtidos na fase de extração de atributos que será descrita na próxima seção.

#### 4.3.6.2 Extração de atributos

Utilizando as amostras de dígitos resultantes do método de segmentação, em conjunto com o arquivo texto contendo o rótulo de classificação manual de cada dígito, criou-se uma base de dígitos rotulados (BDR). Esta base é composta de dígitos padronizados no tamanho de  $40 \times 40$  pixels, que são resultantes do método de segmentação individual de dígitos apresentado neste trabalho. A base é composta de 1220 amostras de dígitos rotulados, com 122 amostras de cada dígito (0 a 9).

Cada amostra apresenta 1722 características dispostas em uma matriz cuja primeira coluna de cada linha é identificada com o rótulo do dígito correspondente (0 a 9). As colunas restantes são compostas por valores extraídos conforme a



aplicação das técnicas indicadas na Tabela 3. Esta tabela apresenta também a quantidade de características extraídas por cada uma das técnicas, bem como o posicionamento dos valores no vetor de atributos.

A duas primeiras técnicas indicadas na tabela, são aplicadas sobre a imagem do dígito em tons de cinza, enquanto que as restantes são aplicadas na imagem limiarizada do dígito.

**Tabela 3. Mapeamento de características extraídas dos dígitos.**

<b>Técnica</b>	<b>Características</b>	<b>Posição inicial no vetor</b>	<b>Posição final no vetor</b>
Mapa de <i>pixels</i>	1600	0002	1601
Vetor de comportamento	16	1602	1617
Zoning 4x4	16	1618	1633
Zoning 4x1 (linhas)	4	1634	1637
Zoning 1x4 (colunas)	4	1638	1641
Zoning diagonal	2	1642	1643
Projeção horizontal	40	1644	1683
Projeção Vertical	40	1684	1723

#### 4.3.6.3 Classificação dos dígitos

O reconhecimento dos dígitos ficou a cargo de três classificadores:

- K-vizinhos mais próximos (*KNN – K-Nearest Neighbours*), descrito na seção 2.8.1 deste documento.
- *Support Vector Machine* (SVM), descrito na seção 2.8.2 deste documento.
- *Extreme Learning Machine* (ELM), descrito na seção 2.8.4 deste documento.

Para o classificador KNN foi escolhido  $k=3$  após avaliação da taxa de acerto em experimentos com  $k$  variando de 1 a 18 (em intervalos de 2 em 2). Para o cálculo da distância foi utilizada a distância Euclidiana.

Para o classificador SVM foi escolhido o *kernel* linear com parâmetro de penalidade  $C=1$ , definidos de forma empírica com base na análise da taxa de acerto no momento do treinamento.

Para o classificador ELM foi escolhida uma arquitetura de 1722 unidades de entrada, que corresponde ao tamanho do vetor de atributos. Na camada oculta foi definida a quantidade de 80 neurônios e 10 sinais de saída (correspondentes aos dígitos de 0 a 9), a função de ativação utilizada foi a sigmoide.

O número de neurônios na camada oculta e a função de ativação foram selecionados empiricamente com base na avaliação da taxa de acerto no momento do treinamento.

Com relação à avaliação do desempenho dos classificadores, foi utilizado o método de validação cruzada *K-fold*, atribuindo a  $k$  o valor de 10 iterações. Foi também estabelecido a divisão de 70% da BDR para aplicação do treinamento e 30% para efeito de teste. A taxa de avaliação foi calculada com base na taxa média dos acertos obtidos em cada iteração.

#### 4.4 CONSIDERAÇÕES FINAIS

Os algoritmos elaborados para o experimento foram desenvolvidos em linguagem de programação *Python 2.7*, utilizando as bibliotecas *PIL* e *Scikit-Image* para o processamento de imagens e a biblioteca *Scikit-Learn* para os classificadores *KNN* e *SVM*, todas disponíveis mediante instalação do pacote de distribuição *Anaconda*<sup>2</sup>.

---

<sup>2</sup> Pacote de distribuição *Anaconda* disponível em: <https://www.continuum.io/downloads> (acessado em 01/06/2016).

Para o classificador ELM foi utilizada a biblioteca *HP-ELM Toolbox*<sup>3</sup>. Os algoritmos foram testados com o apoio de um *notebook* configurado com sistema operacional *Windows 7*, 32bits, 4GB de memória RAM e CPU Intel Core2Duo T6600 de 2.20GHz.

A base rotulada (BDR) utilizada no experimento, assim como as imagens pertencentes aos grupos G1 e G2, juntamente com os dígitos individuais originados do processo de segmentação, está disponível em repositório *online Github*<sup>4</sup>, a fim de que outros pesquisadores possam fazer uso em pesquisas futuras voltadas ao reconhecimento de dígitos. Os experimentos e respectivos resultados são apresentados e discutidos no próximo capítulo.

---

<sup>3</sup> Repositório *Github* da biblioteca *HP-ELM* disponível em: <https://github.com/akusok/hpelm> (acessado em 01/06/2016).

<sup>4</sup> Banco de imagens criado pelo autor deste trabalho disponível em: <https://github.com/jcgcbw/gas-meter-ocr> (acessado em 01/06/2016).

## 5 RESULTADOS E DISCUSSÃO

### 5.1 INTRODUÇÃO

Neste capítulo, são expostas as análises dos resultados, bem como, as considerações a respeito do método proposto e desenvolvido. São demonstrados os resultados obtidos com a aplicação do método sobre as imagens dos grupos G1 e G2, bem como de um terceiro grupo de imagens que não fizeram parte dos dois grupos anteriores.

### 5.2 RESULTADOS OBTIDOS

O experimento do método proposto para localização da ROI, com base no espaço de cores da imagem, foi executado nos grupos de imagem G1 e G2. A Tabela 4 apresenta as taxas de acerto relacionadas ao método de localização da ROI em cada um dos grupos de imagens de medidores.

**Tabela 4. Taxa de acertos na localização da ROI.**

<b>Grupo</b>	<b>Nº de imagens</b>	<b>ROI Localizada</b>	<b>ROI não Localizada</b>	<b>Acerto</b>
G1	640	615	25	96,09%
G2	153	135	18	88,24%

Em uma avaliação visual qualitativa, observou-se que, em média, aproximadamente 92% das 793 imagens processadas tiveram o resultado da detecção da ROI apresentando o contador de consumo de forma integral. Aproximadamente 8% das imagens apresentaram a segmentação da ROI de forma inadequada. Foi possível observar que as falhas em sua maioria foram decorrentes de problemas com a iluminação do ambiente no qual a imagem do medidor foi capturada, bem como, em alguns casos, pela falta da porção vermelha pertencente ao contador de consumo.

A Figura 42 ilustra alguns exemplos de imagens que resultaram em falha na segmentação da ROI. A figura apresenta a imagem do medidor com o respectivo resultado da limiarização para fim de localização de *pixels* vermelhos.



Figura 42. Exemplos de falha na localização da ROI.

Com relação ao método para classificação e reconhecimento dos dígitos, executaram-se experimentos em 1220 amostras de dígitos. Estas amostras são resultantes do método de segmentação individual de dígitos aplicado sobre a ROI segmentada das imagens dos grupos G1 e G2.

Foram utilizados os classificadores ELM, SVM e KNN para o reconhecimento dos dígitos. Os métodos de classificação foram primeiramente aplicados a cada um dos grupos de imagem individualmente, e posteriormente aplicados aos dois grupos em conjunto.

A fim de aferir o desempenho dos classificadores, foi utilizada a técnica de validação cruzada *K-fold* em 10 etapas, observando-se a divisão de 70% dos dígitos para treino do classificador e 30% para teste. A avaliação foi feita com base no acerto médio e no tempo médio de execução.

A Tabela 5 apresenta as taxas de acerto apresentadas em cada uma das etapas da validação cruzada, bem como a taxa média de acertos e o tempo médio decorrido em segundos, com base nos valores retornados pelas 10 etapas de validação. Os resultados são apresentados para cada um dos classificadores ELM, SVM e KNN utilizados no experimento. Para cada classificador é apresentado o resultado da classificação levando em consideração individualmente o grupo de imagens G1 e G2, bem como os dois grupos de imagens em conjunto.

Tabela 5. Taxas de reconhecimento obtidas com o experimento.

Validação Cruzada	Classificador ELM			Classificador SVM			Classificador KNN		
	G1	G2	G1+G2	G1	G2	G1+G2	G1	G2	G1+G2
<i>Fold-1</i>	94%	94%	94%	96%	90%	96%	94%	87%	91%
<i>Fold-2</i>	94%	93%	92%	98%	87%	99%	89%	80%	93%
<i>Fold-3</i>	94%	93%	94%	95%	97%	97%	81%	97%	89%
<i>Fold-4</i>	94%	94%	94%	94%	97%	94%	93%	83%	86%
<i>Fold-5</i>	94%	93%	94%	91%	97%	94%	91%	83%	93%
<i>Fold-6</i>	95%	93%	94%	97%	90%	94%	94%	87%	95%
<i>Fold-7</i>	94%	93%	94%	94%	93%	97%	88%	90%	92%
<i>Fold-8</i>	95%	93%	94%	98%	99%	98%	93%	90%	90%
<i>Fold-9</i>	95%	94%	95%	98%	90%	95%	86%	97%	92%
<i>Fold-10</i>	95%	93%	94%	98%	95%	95%	90%	90%	93%
Acerto Médio	94,4%	93,3%	93,9%	95,9%	93,5%	95,9%	89,9%	88,4%	91,4%
Tempo Médio	0,11 s	0,11 s	0,17 s	1,20 s	0,23 s	1,96 s	4,97 s	0,63 s	8,74 s

Os resultados demonstram classificadores com taxas médias de acerto semelhantes, com o KNN apresentando desempenho ligeiramente inferior em comparação aos métodos ELM e SVM, porém com um alto custo computacional em relação a estes dois últimos. Já a classificação por ELM se destaca por proporcionar um custo computacional bem menor comparado aos demais classificadores.

A fim de validar o método proposto, foi utilizado um novo grupo de imagens composto por 110 imagens de medidores. Assim como no grupo G1, estas imagens também foram fornecidas pela COMPAGAS. Algumas amostras destas imagens podem ser vistas na Figura 43.



**Figura 43. Amostras de imagens do grupo utilizado para validar o método.**

Com relação à fase de segmentação da ROI, para o grupo de imagens de validação o resultado da segmentação foi de 100%. A fase de segmentação individual dos dígitos produziu 544 dígitos que foram submetidos aos classificadores ELM e SVM na fase de reconhecimento.

Devido ao desempenho inferior do classificador KNN em relação ao tempo de processamento, decidiu-se por utilizar apenas os classificadores ELM e SVM na validação do método. A validação levou em conta o percentual de acerto no reconhecimento de dígitos, bem como o percentual no qual o reconhecimento resultou em leitura correta, parcial ou incorreta.

Entenda-se com leitura correta aquela que ocorre quando o resultado do reconhecimento dos dígitos coincide na forma e sequência exata em que a leitura se apresenta na imagem do medidor. Para o diagnóstico de leitura parcial, leva-se em consideração um erro de tolerância, que admite a possibilidade do dígito mais a direita da leitura original estar em desacordo com o resultado do reconhecimento. Qualquer outro tipo de resultado retornado leva ao entendimento de ocorrência de leitura incorreta. A Figura 44 ilustra a forma na qual a classificação da leitura pode ser identificada.

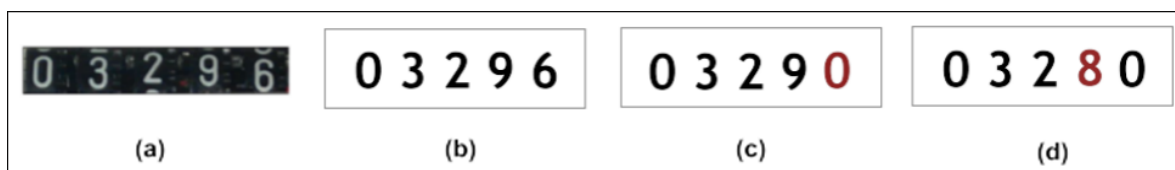


Figura 44. Identificação de leitura. (a) Correta; (b) Parcial; (c) Incorreta.

A Tabela 6 apresenta, para cada classificador, os resultados que correspondem ao diagnóstico de precisão nas leituras, bem como do reconhecimento de dígitos. Juntamente com cada resultado percentual, aparece entre parêntesis o número de leituras ocorridas de forma correta, parcial ou incorreta, bem como quantidade de dígitos reconhecidos corretamente.

Tabela 6. Percentuais de diagnóstico de leitura e reconhecimento de dígitos.

Classificador	Leituras Corretas	Leituras Parciais	Leituras Incorretas	Dígitos Reconhecidos
SVM	80,91% (89)	1,82% (2)	17,27% (19)	94,30% (513)
ELM	84,55% (93)	4,55% (5)	10,91% (12)	95,04% (517)

A Figura 45 apresenta a matriz de confusão gerada com base no resultado obtido com a fase de reconhecimento de dígitos utilizando os dois classificadores.

	0	1	2	3	4	5	6	7	8	9
0	30	3	0	0	0	0	2	0	12	0
1	0	33	0	3	2	0	0	0	0	0
2	0	1	42	0	0	0	0	0	0	0
3	0	0	0	27	0	2	0	0	3	0
4	0	0	0	1	20	0	1	0	0	0
5	0	0	0	0	0	21	0	0	0	0
6	0	0	0	0	0	0	21	0	0	0
7	0	0	0	0	0	0	0	22	0	0
8	0	0	0	0	0	0	0	0	18	0
9	0	1	0	0	0	0	0	0	0	23

(a)

	0	1	2	3	4	5	6	7	8	9
0	40	2	0	2	0	1	0	1	0	1
1	0	35	0	1	1	0	1	0	0	0
2	0	1	41	0	0	0	0	0	1	0
3	0	0	0	29	0	1	1	0	1	0
4	1	0	0	1	20	0	0	0	0	0
5	0	0	0	2	0	17	1	0	0	1
6	0	0	0	0	0	2	18	0	1	0
7	0	0	0	0	0	0	0	22	0	0
8	1	0	0	0	0	0	0	0	17	0
9	1	1	0	0	0	0	0	0	0	22

(b)

Figura 45. Matriz de confusão. (a) Resultado utilizando SVM; (b) Resultado utilizando ELM.

No exercício de validação também foi analisado o tempo médio de execução em cada uma das fases do método proposto. Foi investigado o tempo médio decorrido em segundos para segmentar a ROI, segmentar os dígitos e extrair suas características, bem como executar o reconhecimento dos dígitos tanto utilizando o



classificador SVM quanto o classificador ELM. Foi observado também o tempo total decorrido para execução completa do método utilizando cada um dos classificadores. A Tabela 7 apresenta os resultados obtidos.

**Tabela 7. Tempo médio de execução das fases do método.**

Segmentação ROI	Segmentação Dígitos	OCR via SVM	OCR via ELM	Total com SVM	Total com ELM
0,73 s	0,21 s	2,13 s	0,24 s	3,07 s	1,18 s

## 5.5 DISCUSSÕES GERAIS

Com as taxas de acerto obtidas no experimento foi possível confrontar os resultados com alguns casos apresentados na literatura nos quais os autores destacaram as taxas de acertos relacionadas ao processo de reconhecimento dos dígitos. A Tabela 8 destaca, de acordo com o trabalho de cada autor, as seguintes informações: a taxa de acerto e o tipo de classificação utilizada para o reconhecimento dos dígitos, a presença de controle na captura das imagens (câmera ou outro dispositivo de captura acoplado ao medidor) e o número de imagens de medidores utilizadas no experimento.

**Tabela 8. Taxas de reconhecimento obtidas nos trabalhos relacionados.**

Autor	Controle	Classificado por	Acerto	Imagens
Castells-Rufas e Carrabina (2006)	Sim	Programação Dinâmica	99%	396
Zhao et al. (2009)	Sim	RNA (algoritmo BP)	95%	500
Bai et al. (2010)	Sim	Momentos Hu	94%	50
Lei et al. (2010)	Sem indicação	Análise esqueleto	99%	850
Nava-Ortiz, Gómez e Diaz-Perez (2011)	Sim	Distância mínima	93%	217
Nodari e Gallo (2011)	Não	OCR Tesseract	58%	153

Jabba et al. (2012)	Sim	RNA (algoritmo BP)	90%	180
Zhao et al. (2013)	Sem indicação	RNA (algoritmo BP)	99%	100
Método empregado neste trabalho	Não	ELM	95%	903

Analisando-se os resultados apresentados na literatura, foi possível concluir que o resultado obtido neste trabalho de aproximadamente 95% de taxa de acerto no reconhecimento dos dígitos, tanto no melhor resultado do *k-fold*, quanto no experimento de validação, pode ser considerado promissor. Vale ressaltar que em boa parte dos casos levantados na literatura as imagens de medidores são capturadas em ambientes controlados ou em ambiente de laboratório, ao passo que neste trabalho as imagens foram capturadas em seu ambiente real.

Dentre os trabalhos destacados na Tabela 8, é possível observar taxas de acerto superiores ao resultado obtido utilizando o método apresentado neste trabalho. Porém, no caso do trabalho de Castells-Rufas e Carrabina (2006), o ambiente é controlado por meio de um sensor CMOS em conjunto com um dispositivo FPGA que são acoplados ao medidor para captura das imagens. No trabalho de Jabba et al. (2012), o autor trabalha com imagens de um modelo específico de medidor de energia elétrica, também capturadas em laboratório. Relatando um nível de confiança no reconhecimento de 90% chegando a 100% em determinados dígitos.

No caso do trabalho de Zhao et al. (2013), a imagem não é relacionada diretamente ao medidor, mas apenas ao ciclômetro que é a peça responsável por fazer a contagem do volume conforme pode ser visto na figura 46. Embora o trabalho não faça referência ao controle de ambiente, é possível presumir que a captura das imagens foi feita em laboratório, já que a presença apenas do ciclômetro não é característica do cenário natural no qual se apresentam os medidores de gás.



**Figura 46 - Ciclômetro que realiza a contagem do volume de gás.**

A impossibilidade de acesso às imagens utilizadas por cada um dos pesquisadores em seus trabalhos dificulta a tarefa de comparação do resultado obtido nesta pesquisa com os trabalhos já publicados. Mesmo que se possa ter uma presunção de que imagens capturadas em ambiente controlado favoreçam o alcance de taxas altas de reconhecimento dos dígitos, a aplicação do método proposto utilizando as imagens manuseadas nos trabalhos já publicados resultaria em uma comparação mais efetiva. Há por exemplo o caso do trabalho de [Lei et al. 2010], que além de não terem sido disponibilizadas as imagens, não foi destacado se o ambiente de captura das imagens possuía ou não algum tipo de controle.

Outro meio seria o caminho inverso, reproduzir os métodos empregados nos trabalhos publicados, utilizando as imagens manuseadas nesta pesquisa, porém, isto acarretaria em um acréscimo de tempo que poderia ser incompatível com o tempo determinado para conclusão desta pesquisa.

Dentre os trabalhos da literatura, a comparação mais pertinente seria com a pesquisa de Nodari e Gallo (2011), que disponibiliza a base de imagens, e na qual o autor obteve resultado de 58% no reconhecimento dos dígitos e que, assim como neste trabalho, utiliza imagens capturadas em campo e que apresentam maior conformidade com a situação encontrada pelos profissionais de leitura no seu dia a dia. Uma parcela do grupo de imagens utilizadas pelo autor, compôs o grupo de imagens G2 utilizado neste trabalho.

Com relação ao tempo de execução, em apenas três casos os autores destacaram tal informação, mesmo que de forma superficial. Nos três casos a classificação foi efetuada por meio da aplicação de uma RNA utilizando o algoritmo BP.

Em Zhao et al. (2009) o autor descreve um tempo de 9 s para a realização do método, não fornecendo informações mais detalhadas com relação ao tempo de treinamento da RNA ou mesmo o tempo para a fase de segmentação.

No trabalho de Jabba et al. (2012), são utilizadas cinco combinações distintas de RNAs, para as quais o tempo de treinamento máximo obtido foi próximo de 255 s, e o tempo mínimo próximo de 12 s. O tempo de reconhecimento de cada dígito foi próximo de 6 ms. Não foi destacado o tempo para a fase de segmentação, bem como para execução do método como um todo.

Por fim, a pesquisa de Zhao et al. (2013) não destacou o tempo empregado na fase de segmentação. O tempo de reconhecimento de cada dígito foi em média 24 ms, não destacando o tempo para execução do treinamento da rede.

A análise efetuada neste trabalho apresentou o custo identificado em cada fase do método, desde a segmentação da ROI, segmentação individual dos dígitos, reconhecimento e custo total. Desta forma não foi possível efetuar uma comparação abrangente relacionada ao custo computacional para execução do método.

## 6 CONCLUSÕES E PERSPECTIVAS PARA TRABALHOS FUTUROS

### 6.1 CONCLUSÕES E DISCUSSÕES

Este trabalho teve como objetivo apresentar a metodologia adotada para a construção de um algoritmo envolvendo técnicas de processamento de imagem e reconhecimento de padrões. O algoritmo visa obter em imagens de medidores de gás natural, o reconhecimento dos dígitos apresentados no medidor com o propósito de aprimorar o processo de coleta de consumo em distribuidoras de gás natural.

A relevância desta pesquisa se dá, do ponto de vista tecnológico e industrial, pela aplicabilidade em um problema atual apresentado em ambiente real das distribuidoras de gás natural. Tal pesquisa vai ao encontro de um dos objetivos do PPGCA<sup>5</sup> que é o de gerar conhecimento tecnológico e científico para resolver problemas regionais e nacionais, bem como permitir a transferência tecnológica entre Universidade e empresas. Do ponto de vista científico, a proposta busca contribuir com o avanço do estado-da-arte de detecção e reconhecimento de dígitos.

O método aplicado neste trabalho teve como propósito a utilização de imagens de medidores capturadas em situações diversas, mais próximas do ambiente real encontrado pelos profissionais leiturista em seu dia-a-dia.

Para segmentar a ROI contendo o contador de consumo localizado no medidor, partiu-se da hipótese de utilização do espaço de cor (HSL) da imagem, pressupondo que a presença da cor vermelha apresentada na ROI contribuísse no processo de localização. O método de segmentação foi aplicado em 903 imagens pertencentes aos três grupos de imagens utilizados na pesquisa. A taxa de acerto na localização da ROI foi, em média, próxima de 95%. As falhas ocorreram principalmente por problemas de iluminação do ambiente, e em casos nos quais a porção vermelha do contador de consumo se mostrava de forma insuficiente.

Assim, a hipótese de segmentar a ROI com base em um método que utilizasse o espaço de cores HSL da imagem mostrou-se acertada, pois dentre as imagens utilizadas no experimento, em menos de 6% dos casos a segmentação da ROI apresentou resultado que impossibilitasse a identificação dos dígitos.

---

<sup>5</sup> Objetivos do PPGCA disponível em: <http://ppgca.dainf.ct.utfpr.edu.br/doku.php?id=descricao> (acessado em 19/07/2016).

A continuação da aplicação do método de segmentação, a fim de localizar individualmente os dígitos, gerou 1220 amostras de dígitos que foram utilizadas para a comparação do desempenho dos algoritmos de classificação KNN, SVM e ELM. Assim como gerou 544 amostras de dígitos reservadas para validação do método.

Para aferir o desempenho dos classificadores, foi utilizada a técnica de validação cruzada *K-fold* em 10 etapas, com a separação de 70% das amostras para treino dos algoritmos e 30% para teste. Os resultados demonstraram classificadores com desempenhos semelhantes, com o KNN tendo um desempenho ligeiramente inferior em comparação aos métodos ELM e SVM, e apresentando alto custo computacional. Já a classificação utilizando ELM teve um desempenho muito semelhante ao SVM, mas com um custo computacional bem menor.

Como o KNN apresentou alto custo computacional, foram validados apenas os classificadores SVM e ELM, utilizando as 544 amostras de validação. O resultado reforçou o que foi apresentado pelo método *K-fold*, com taxas de acerto semelhantes entre os classificadores, e tempo de processamento consideravelmente menor apresentado pelo método ELM. O classificador ELM reconheceu 95,04% dos 544 dígitos apresentados, enquanto que o classificador SVM reconheceu 94,3% dos dígitos. Para o reconhecimento do valor de consumo apresentado no medidor, o classificador ELM foi também superior apresentando 10,91% de leituras incorretas, enquanto que o classificador SVM apresentou 17,27% de leituras incorretas.

Em relação aos resultados apresentados pela literatura, o método proposto baseado em ELM, igualou ou superou o desempenho em metade dos casos apresentados, ficando próximo dos resultados superiores. Ressalte-se que, na maioria dos trabalhos a captura da imagem é efetuada com controle no ambiente.

Com relação ao custo computacional para reconhecer o valor de consumo apresentado no medidor, o método ELM obteve um tempo de processamento de 0,24 s por imagem. O SVM apresentou um tempo de 2,13 s por imagem. Este resultado confirmou a expectativa de que o classificador ELM poderia apresentar desempenho melhor em relação a métodos tradicionais de classificação, devido a sua característica de redução significativa no tempo de treinamento. .

Outro ponto a ser destacado com relação ao custo computacional é o percentual de tempo observado em cada fase do método. A fase com maior tempo de processamento foi a de segmentação da ROI, responsável por 62% do tempo utilizado pelo método. A fase de reconhecimento foi responsável por 20%, enquanto que a fase de segmentação individual dos dígitos foi responsável por 18% do tempo utilizado. O tempo total de execução do método foi em média de 1,18 s por imagem.

Em relação aos trabalhos da literatura, em apenas três casos foram destacados resultados, porém em nenhum destes observou-se o tempo decorrido de modo separado em cada fase do método. Desta forma não foi possível efetuar uma comparação abrangente relacionada ao tempo de execução

De maneira geral o método proposto se apresenta promissor e atendeu ao objetivo desta pesquisa, já que com a utilização do classificador ELM foram alcançadas taxas de acerto em torno de 95% com baixo custo computacional. Apresentou-se robusto, já que foram utilizadas 903 imagens de medidores obtidas em campo, apresentando menos de 6% de falha no processo de segmentação da ROI. O tempo total de execução apresentado pelo método, em média 1,18 s, se mostra adequado para a execução em uma aplicação em tempo real utilizando dispositivos móveis. Porém, é possível presumir que incorporando determinado nível de controle na captura das imagens possa ser promovida uma melhoria no processo. Um exemplo de controle seria, no momento da captura da imagem, apresentar em tela um retângulo englobante no qual a área da ROI seria encaixada. A aplicação poderia então efetuar uma pré-segmentação da ROI, diminuindo assim o espaço de busca utilizado pela rotina que segmentará o contador de consumo. Como a segmentação é a fase mais custosa do método, pois utiliza a área total da imagem para localizar a ROI, a utilização apenas da área pré-segmentada abreviaria o tempo gasto para segmentar o contador de consumo.

As principais contribuições deste trabalho são:

- Em face da escassez de bases de dados compatíveis com o objetivo deste trabalho, é disponibilizada online<sup>6</sup> uma base contendo 903

---

<sup>6</sup> <https://github.com/jcgcbw/gas-meter-ocr>.

imagens de medidores, composta por imagens de tamanhos variados, obtidas em campo a partir de condições reais de operação.

- Um método que efetue a segmentação do contador de consumo apresentado em imagens coloridas de medidores de gás, obtidas em campo a partir de condições reais de operação (diferentemente da maioria das abordagens da literatura, que tomam como base imagens em tons de cinza, e com a captura efetuada em laboratório).
- A utilização de Máquinas de Aprendizagem Extrema - (ELM), um método de classificação que apresenta como diferencial uma redução significativa no tempo de treinamento, permitindo sua utilização em aplicações em tempo real. Este método não foi abordado nos trabalhos encontrados na literatura com o mesmo objeto de estudo.
- Um estudo comparativo entre os classificadores KNN, SVM e ELM, aplicados na classificação de dígitos apresentados em medidores de gás natural. O estudo demonstra a efetividade de ELM no que diz respeito ao custo computacional empregado para classificação.

Alguns cenários para a aplicabilidade desta pesquisa poderiam ser:

- Disponibilizar ao profissional leitorista um aplicativo em dispositivo móvel como forma alternativa de coleta de dados de consumo.
- Disponibilizar ao cliente o mesmo aplicativo, caso este não deseje agendar a visita do profissional leitorista para efetuar a coleta.
- Aplicação em processamento *batch* para verificação de leituras coletadas por meio de digitação e com foto capturada.
- Aplicação para auxílio em laboratórios de calibração de medidores.

As principais limitações do método proposto são relacionadas ao reconhecimento de dígitos em transição, ser afetado por variação na escala dos dígitos e dificuldade de localizar a ROI em imagens com pouca luminosidade. Estas limitações serão abordadas em trabalhos futuros relacionados na próxima seção.

## 6.2 TRABALHOS FUTUROS

Como sugestão de trabalhos futuros recomenda-se a utilização de outros descritores de características a fim de lidar com as limitações relacionadas na seção



anterior. Em relação à variação de escala sugere-se a utilização dos momentos invariantes de Hu. Já características estruturais como a presença de concavidades, furos e traços poderia ser um caminho para os casos de dígitos em transição.

Ainda com relação aos dígitos em transição, seria útil também a captura de mais imagens nas quais os dígitos se apresentem neste estado. Com relação a falta de luminosidade um caminho seria o emprego de um filtro para ajuste de contraste.

Seria também interessante experimentar o algoritmo com um hardware de baixo custo como um dispositivo celular, possibilitando assim o levantamento da quantidade de leituras em campo obtidas de forma acertada.

Além disto, sugere-se a aplicação de um moderado nível de controle na captura das imagens, por exemplo, fazendo com que no momento da captura, o contador de consumo esteja posicionado no interior de um aparente retângulo englobante presente na tela do dispositivo. Embora vá de encontro à proposta inicial, a incorporação de um nível de controle na captura poderia trazer benefícios tanto no que diz respeito ao problema de escala, quanto na diminuição de tempo de processamento, pois o espaço de busca para localização da ROI seria diminuído. Uma elevação da taxa de reconhecimento também poderia ser observada.

## REFERÊNCIAS

ACHARYA, Tinku; RAY, Ajoy K. **Image processing: principles and applications**. John Wiley & Sons, 2005.

ALEXANDRIA, Auzuir R. **Sistema de Reconhecimento de Algarismos em Medidores de Energia Convencionais Visando Implementação de Sistema de Fornecimento de Energia Pré-Paga – Seppra**. 2004. Dissertação de Mestrado, Universidade Federal do Ceará, 2004.

ARRIGHI, T.; ROJAS, J. E.; SOTO, J. C.; MADRIGAL, C. A.; LONDOÑO, J. A. Recognition and classification of numerical labels using digital image processing techniques. In: **2012 XVII Symposium of Image, Signal Processing, and Artificial Vision (STSIVA)**. IEEE, 2012. p. 252-260.

BAI, Qiushi; ZHAO, Limeng; ZHANG, Yunzhou; QI, Zixin. Research of automatic recognition of digital meter reading based on intelligent image processing. In: **Computer Engineering and Technology (ICCET), 2010 2nd International Conference on**. IEEE, 2010. p. V5-619-V5-623.

BELAN, P. A.; ARAUJO, S. A.; LIBRANTZ, A. F. H. Segmentation-free approaches of computer vision for automatic calibration of digital and analog instruments. **Measurement**, v. 46, n. 1, p. 177-184, 2013.

BERNSEN, John. Dynamic Thresholding of gray-level images, In: **Proc. Eighth Int'l Conf. on Pattern Recognition**, 1986. Paris, France.

BOARETTO, Neury. **Detecção e Classificação Automática de Defeitos de Soldagem em Tubulações: Uma Abordagem com Ensemble de Redes Neurais Extreme Learning Machines**. 2016. Qualificação de Doutorado, Universidade Tecnológica Federal do Paraná, 2016.

EDWARD, Cephas P. Support Vector Machine based automatic electric meter reading system. In: **Computational Intelligence and Computing Research (ICCIC), 2013 IEEE International Conference on**. IEEE, 2013. p. 1-5.

EPSHTEIN, Boris; OFEK, Eyal; WEXLER, Yonatan. Detecting text in natural scenes with stroke width transform. In: **Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on**. IEEE, 2010. p. 2963-2970.

CASTELLS-RUFAS, David; CARRABINA, Jordi. Camera-based digit recognition system. In: **2006 13th IEEE International Conference on Electronics, Circuits and Systems**. IEEE, 2006. p. 756-759.

DOS SANTOS, Eulanda M. **Teoria e aplicação de support vector machines à aprendizagem e reconhecimento de objetos baseado na aparência**. 2002. Tese de Doutorado. Universidade Federal da Paraíba.

DOUGHERTY, Edward R. An introduction to morphological image processing. **Tutorial texts in optical engineering**, 1992.

GAZCÓN, Nicolás F.; CHESÑEVAR, Carlos I.; CASTRO, Silvia M. Automatic vehicle identification for Argentinean license plates using intelligent template matching. **Pattern Recognition Letters**, v. 33, n. 9, p. 1066-1074, 2012.

GONZALEZ, Rafael C.; WOODS, Richard E. **Digital Image Processing**. 2a. ed. EUA: Addison-Wesley Publishing Company, 2002. 793p.

HAMERSCHMIDT, Marcio B. **Telemedições aplicadas ao faturamento de energia elétrica: um diagnóstico de viabilidade técnica e econômica na realidade brasileira**. 2012. 126 f. Dissertação de Mestrado, Lactec, Curitiba, 2012.

HU, Ming-Kuei. Visual pattern recognition by moment invariants. **IRE transactions on information theory**, v. 8, n. 2, p. 179-187, 1961.

HUANG, Guang-Bin; ZHU, Qin-Yu; SIEW, Chee-Kheong. Extreme learning machine: a new learning scheme of feedforward neural networks. In: **Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on**. IEEE, 2004. p. 985-990.

HUANG, Guang-Bin; ZHU, Qin-Yu; SIEW, Chee-Kheong. Extreme learning machine: theory and applications. **Neurocomputing**, v. 70, n. 1, p. 489-501, 2006.

HUANG, Guang-Bin; WANG, Dian H.; LAN, Yuan. Extreme learning machines: a survey. **International Journal of Machine Learning and Cybernetics**, v. 2, n. 2, p. 107-122, 2011.

HOUGH, Paul VC. **Method and means for recognizing complex patterns**. 1962.

IBRAHEEM, Noor A.; HASAN, Mokhtar M.; KHAN, Rafiqul Z.; MISHRA, Pramod K. Understanding color models: a review. **ARPN Journal of Science and Technology**, v. 2, n. 3, p. 265-275, 2012.

JABBA, Daladier; RODRIGUEZ, Miguel; BERDUGO, Geovanni; CALLE, Maria; JIMENO, Miguel A.; ZUREK, Eduardo E. A pattern recognition procedure for the identification of digital numbers in electrical meters using Neural Networks. In: **2012 IEEE Symposium on Industrial Electronics and Applications**. 2012.

JUNG, Keechul; KIM, Kwang In; JAIN, Anil K. Text information extraction in images and video: a survey. **Pattern recognition**, v. 37, n. 5, p. 977-997, 2004.

LEI, Haijun; ZHANG, Panpan; ZENG, Qingqian; LI, Xianyi. Numeral recognition of power meter on a handheld terminal. In: **Proceedings of the Third International Symposium on Electronic Commerce and Security Workshops**. 2010. p. 76-79.

LIM, Junsik; KIM, Soohyung; PARK, Jong-Hyun; LEE, Guee-Sang; YANG, HyungJeong; LEE, Chil-Woo. Recognition of Text in Wine Label Images. In: **Pattern Recognition, 2009. CCPR 2009. Chinese Conference on**. IEEE, 2009. p. 1-5.

LUNTS, A. L.; BRAILOVSKY, V. L. Evaluation of attributes obtained in statistical decision rules. **Engineering Cybernetics**, n. 3, p. 98-&, 1967.

MARQUES FILHO, Oge; VIEIRA NETO, Hugo. **Processamento digital de imagens**. Brasport, 1999.

NAVA-ORTIZ, M.; GÓMEZ, Wilfrido; DIAZ-PEREZ, A. Digit recognition system for camera mobile phones. In: **Electrical Engineering Computing Science and Automatic Control (CCE), 2011 8th International Conference on**. IEEE, 2011. p. 1-5.

NEEDLEMAN, Saul B.; WUNSCH, Christian D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. **Journal of molecular biology**, v. 48, n. 3, p. 443-453, 1970.

NIBLACK, Wayne, (1986). **An Introduction to Digital Image Processing**. Englewood Cliffs, N. J., Prentice Hall, 1986. 115-116.

NODARI, Angelo; GALLO, Ignazio. A Multi-Neural Network Approach to Image Detection and Segmentation of Gas Meter Counter. In: **12th IAPR Conference on Machine Vision Applications**. 2011. p. 239-242.

OBINATA, Goro; DUTTA, Ashish. **Vision Systems: Segmentation and Pattern Recognition**. I-Tech Education and Publishing, 2007.

OLIVEIRA, Danilo M.; CRUZ, Ruironald S.; BENSEBAA, Kamel. Automatic numeric characters recognition of kilowatt-hour meter. In: **Signal-Image Technology & Internet-Based Systems (SITIS), 2009 Fifth International Conference on**. IEEE, 2009. p. 107-111.

OTSU, Nobuyuki. (1979), A Threshold Selection Method from Gray-Level Histograms, **IEEE Transactions on Systems, Man and Cybernetics**, p. 62-66.

PEDRINI, Hélio; SCHWARTZ, William R. **Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações**, Thomson, São Paulo, 2008. 508p.

PITAS, Ioannis. **Digital Image Processing Algorithms and Applications**. EUA: John Wiley and Sons, 2000. 432p.

PRATT, William K. **Digital Image Processing PIKS Scientific Inside**. 4.ed. Los Altos, California, USA: John Wiley & Sons, Inc., 2007. 769p.

DE QUEIROZ, José E. R.; GOMES, Herman M. Introdução ao Processamento Digital de Imagens. **Revista de Informática Teórica e Aplicada**, v. 13, n. 2, p. 11-42, 2006.

RUDEK, Marcelo. **Uma proposta para indexação e recuperação automática de imagens e reconhecimento de cheques bancários baseadas no vetor de comportamento**. 1999. Dissertação de Mestrado, CPGEI CEFET-PR. Curitiba.

SERRE, D. (2002). **Matrices: Theory and Applications**. Springer, 1st edition.

SILVESTRE, Leonardo J. **Regularização de Extreme Learning Machines: uma abordagem com matrizes de afinidade**, Tese de Doutorado, UFMG, Belo Horizonte, 2015.

SHU, Dongmei; MA, Shuhua; JING, Chunguo. Study of the automatic reading of watt meter based on image processing technology. In: **2007 2nd IEEE Conference on Industrial Electronics and Applications**. IEEE, 2007. p. 2214-2217.

SMITH, Ray. An Overview of the Tesseract OCR Engine. In **proceedings of Document analysis and Recognition**. ICDAR 2007. In: IEEE Ninth International Conference.

STONE, Mervyn. Cross-validators: choice and assessment of statistical predictions. **Journal of the Royal Statistical Society. Series B (Methodological)**, p. 111-147, 1974.

SOILLE, Pierre. **Morphological Image Analysis: principles and applications**. 2.ed. Secaucus, NY, USA: Springer-Verlag New York, Inc., 2003. 391p.

VANETTI, Marco; GALLO, Ignazio; NODARI, Angelo. GAS meter reading from real world images using a multi-net system. **Pattern Recognition Letters**, v. 34, n. 5, p. 519-526, 2012.

VAPNIK, Vladimir. **The nature of statistical learning theory**. Springer Science & Business Media, 2013.

WEN, Ying; LU, Yue; YAN, Jingqi; ZHOU, Zhenyu; VON DENEEN, Karen M.; SHI Pengfei. An algorithm for license plate recognition applied to intelligent transportation system. **IEEE Transactions on Intelligent Transportation Systems**, v. 12, n. 3, p. 830-845, 2011.

ZHANG, Jing; CHEN, Xilin; HANNEMAN, Andreas; YANG, Jie; WAIBEL, Alex. A robust approach for recognition of text embedded in natural scenes. In: **Pattern Recognition, 2002. Proceedings. 16th International Conference on**. IEEE, 2002. p. 204-207.

ZHAO, Limeng; ZHANG, Yunzhou; BAI, Qiushi; ZHANG, Xintong. Research of digital meter identifier based on DSP and neural network. In: **Imaging Systems and Techniques**, 2009. IST'09. IEEE International Workshop on. IEEE, p. 402-406.

ZHAO, Wei; GAO, MingYu; HE, Zhi Wei. Character Recognition Based on Improved BP Neural Network. In: **Instrumentation, Measurement, Computer, Communication and Control (IMCCC)**, 2013 Third International Conference on. IEEE. p. 754-757.